

968 -----GlnAsnCyseYsAlaPro---ProAlaPro----- 976

Qy 315 TCAGAGCCAGCTGTATGGCAATTAATTGCTACAGGGGTGGTATCGCTGCATACCGCTACGC 374

Db 977 -----CysCyseYsLeuProThrIle 982

Qy 375 CCAGCCACACCCCTGCGACCTGCGTCCACAGTGCATGACGTTACGAGACAGATTATGCTGC 434

Db 983 ProCyseYsProProIleProCyseYsProGln-----ProYleIleCyseYs 998

Qy 435 CGAGCC-----CTACACACAGACACTTCTCTCAGC----- 464

Db 999 GlnProAlaProValCyseYsLeuProProProThrIleCysYsSerIleAlaSnPhelYsLeuPro 1018

Qy 465 -----CCCGACCTACGGCGCTGTGTCATGATGCTTTTGC 500

Db 1019 ThrIleProIleCyseYsGlyAlaGlnAlaCyseProSerCyseYsProCyseYs 1032

RESULT 7

B89716

protein F45B8.3 [imported] - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Nov-2001

C/Accession: B89716

R:Anonymous, The C. elegans Sequencing Consortium.

A/Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog

A/Note: see websites genome.muscl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.ele

A/Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and

A/Accession: B89716

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-211 <STO>

A/Cross-references: GB:chr_X; PID: CAB05726.1; PID: G3877144; GSPDB: GN00028; CESP: F45B8.3

A/Genetics:

A/Gene: F45B8.3

A/Map position: X

Alignment Scores:

Pred. NO.: 1.09 Length: 211

Score: 100.50 Matches: 46

Percent Similarity: 38.12% Conservative: 15

Best Local Similarity: 28.75% Mismatches: 65

Query Match: 8.63% Indels: 34

DB: 2 Gaps: 6

US-09-809-545A-1_cOPY_535_1143 (1-609) x B89716 (1-211)

Qy 15 GGCGGTGAACCCCTTACACCAATGGCTGGAATTAATCAGTGTGGGGCGGCTACAG 74

Db 56 GlyProGlnProPheAsnProValPheAsnIleHisPheAsnCyseYsGlyAlaProArg 75

Qy 75 CCCGCACTTCTATGC-----AGCAGCGTGTGTTGTCAGAGGCCAACAGAGAGG 125

Db 76 ProSerProSerCyseYsProIleProValProProAlaProLeuProProProPro- 95

Qy 126 ATCTTCATGTACAGTGGCCCGCACTTCACTTGTATATACTTGCAGAAAGCCTGCTTTC 185

Db 95 oAlaSerProCyseYsGlyProSerProVal-----ProAlaProCyseYsPr 111

Qy 166 ATATCCGGCGCGCAGCTGCTGAGCTGACATACCGAGGGGCTCACTTCAGAGCGTGTGCG 245

Db 111 oProProProAlaProAlaProAlaProCys----- 120

Qy 246 CACCGTGTACAAACACCTTCAGAGCTGGCGGCGCCCGCCACCCCAATCCGGGCTTATGGCGG 305

Db 121 -----CysProProProProProProProProProProProSerProLeuVa 133

Qy 306 AGTAGTGTATCAGAGCCAGCTGTATGGCAATTAATTGCTACAGGGGTGTACGCTGCATA 365

Db 133 lCysCyseYsGlnAlaProValProGlnAsnPro-CysCyseYsGlnIleValAla-----A 151

```

QY 366 CCGGTAACGCCAGCCACCC-----CTGCACTGCTGCTGCTACAGTACAGTACGAGC 422
Db 151 LAALMePrProPrProPrProPrProPrProPrProPrProPrProPrProPrProPr 170
QY 423 AGTTATGCTGCGCAGCCCTACCA-----CCACACACTTGCTCCAGCCC 467
Db 171 AenProCyCysGlnProAlaProAlaProAlaProCysValCySserAlaPro 188

RESULT 8
150703
transcription factor GATA-6 - chicken
C.Species: Gallus gallus (chicken)
C.Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #ext_change 26-Aug-1999
C.Accession: 150703
R.Laverriere, A.C.; MacNeill, C.; Mueller, C.; Poelmann, R.E.; Burch, J.B.; Evans, T.
J. Biol. Chem. 269, 23177-23184, 1994
A.Title: GATA-4/5/6, a subfamily of three transcription factors transcribed in developing
A.Reference number: A54720; MUID:94365018; PMID:8083222
A.Accession: 150703
A.Status: translated from GB/EMBL/DBJ
A.Molecule type: mRNA
A.Residues: 1-387 <LA>
C.Comment: This transcriptional activator is named for the core, GATA, of the nucleotide
C.Genes:
A.Gene: GATA-6
C.Superfamily: transcription factor GATA-4, GATA-type zinc finger homology
C.Keywords: DNA binding; nucleus; transcription factor; transcription regulation; zinc fi
F.178-231/Domain: GATA-type zinc finger homology <GF1>
F.181-205/Region: zinc finger GATA motif
F.232-285/Domain: GATA-type zinc finger homology <GF2>
F.235-259/Region: zinc finger GATA motif

Alignment Scores:
Pred. No.: 1 43 Length: 387
Score: 99.00 Matches: 53
Percent Similarity: 35.94% Conservative: 16
Best Local Similarity: 27.60% Mismatches: 50
Query Match: 8.51% Indels: 73
DB: 2 Gaps: 11

US-09-809-545A-1_COPY_535_1143 (1-609) x 150703 (1-387)
QY 27 CTACACCAATGCGTGGAAATTAATCAGTTGGGCGGCTTACAGCCCGACTTCTA 86
Db 1 MeTyxGlnThrLeuAlaIleSerAlaSerGlnGly----- 12
QY 87 TCGAGCAGCGGCTGTTGTGCGAGCCAAACGAGGGAATCTTCATGTACAGTGGCC 146
Db 13 -----ProAlaProTyxArgGlySerProGlyGly--Phe-MethisSeraLPr 28
QY 147 CAGTTCACCTGTATATAC-----TCGCAATGCGCTGCTTCATA 188
Db 28 oSerSerProValTyxValProThrThrArgValGlySerValLeuProThrLeuProTy 48
QY 189 T----- 189
Db 48 rIeuGlnGlyGlyAlaAlaGlnProGlyHisAlaProAlaGlyHisValTpsrG 68
QY 190 -CCGCGCGCACTGCTGAGCT-----GCATCCGAGGGGCTTCCAGCGCGTGG 242
Db 68 nProAlaIaIaGlySerProSerTyxGlyAlaIaIaGlyAlaHisProSerGlyArg-- 87
QY 243 TCGCACCGGTACACACCTTCAGAGCTGGGGGGCCCCCACCCTCATC----- 291
Db 88 -----PheProTyxSerAlaSerProProAlaAlaAsnGlyAla 101
QY 292 ----CCGCGCTATGCGGAGTAGTATCAAGACCGCAGTATAGC-----AATTAATT 341
Db 101 rArgGlnGlnTyxGlyGlyGlyLeuAlaAlaArgGlnGlnTyxGlyAlaLeuProAlaGPr 121
QY 342 GCTACAGGGGTGTTAGCGTCATACCGCTACGCCGACGCCACCCCTGCCACTGCTGTC 401

```

Db 121 oleuansglyserTy-----ProAlaProTyrAlaSe 132
Qy 402 CTAC---AGTGACACTGACGAGATTATGCTCCGACCCCTACACACACTT-- 456
Db 132 rTyValGlyProGlnLeuGlyProAlaTyrProAlaAlaProPheGluAenSerValle 152
Qy 457 -----GCTCCAGCCCC 468
Db 152 uHieCyBeuGlnGlyArgAlaAlaProIlePro 163

RESULT 9
T29707
hypothetical protein ZK180.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C:Accession: T29707
R:Pauley, A.; Le, T.T.
Submitted to the EMBL Data Library, May 1996
A:Description: The sequence of C. elegans cosmid ZK180.
A:Reference number: 220669
A:Accession: T29707
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-542 <PAU>
A:Cross-references: EMBL:U58748; PIDN:AAB52969.1; GSPDB:GN00022; CESP:ZK180.5
C:Experimental source: Strain Bristol N2; clone ZK180
C:Genetics:
A:Gene: CESP:ZK180.5
A:Map position: 4
A:Insertions: 3/3; 47/2; 484/3
A:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

Alignment Scores:
Pred. No.: 1.41 Length: 542
Score: 99.00 Matches: 51
Percent Similarity: 29.18% Conservative: 17
Best Local Similarity: 21.89% Mismatches: 71
Query Match: 8.51% Indels: 94
DB: 2 Gaps: 8

US-09-809-545A-1_COPY_535_1143 (1-609) x T29707 (1-542)

Qy 106 TGCAGGCCCAACGAGAGGA---TCTTCATGTACAGTGGCCCGCTTCACTTGTAT 162
Db 29 CyGlnProSerGlnProAlaCyAlaAenProCyGlyAlaSerAlaAlaTyr 48
Qy 163 ACTTCGCAATGCTGCTTTCATATCCGGCCGACCTGTCAGCTCATACGAGGG 222
Db 49 AlaAlaProAlaAlaGlyTyrProArgProProIle----- 60
Qy 223 GCTCACCTTCGAGCGCTGCGACCGGTACACACCTTGAGAGCTGCGGCCCA 282
Db 61 -----GlyTyrGly-----AlaProProProPro 68
Qy 283 CCCCACATCCGGCTATGCGGAGTA----- 309
Db 69 ProProProProAlaPheGlyGlyLeuGlyGlyAlaAlaTyrProGlyAlaGlyAla 88
Qy 310 -----GTCTATCAAGAGCGAGTATGGCAAT----- 336
Db 89 GlyGlyGlnTyrAspThrGlyValHisGlyGlyAlaGlnGlyGlyTyrAlaGlyAlaGln 108
Qy 337 -----AAATTGCTACAGGGTGTACCTGACGTATCCGC 369
Db 109 GlyAlaGlnGlyGlyTyrAlaGlyAlaGlnGlyAlaGlnGlyGlyTyrAlaGlyAlaGln 128
Qy 370 TACGCCACGCCACCCCTGACCTGCTGCT----- 399
Db 129 GlyAlaGlnGlyGlyTyrAlaGlyAlaAlaGlnGlyAlaGlnGlyGlyTyrAlaGlyAla 148
Qy 400 -----GCCTACAGTGCACAGTTACGAGAGAGTTATGCT 432
Db 149 ValGlnGlyGlyGlnIleAlaSerGlnGlyTyrAlaGlyAlaGlyAlaProValSerAla 168

Qy 433 GCCGACCCCTACCAACAC----- 450
Db 169 GlyIseTyrAenGlnGlyProAlaAlaIleAenGlyTyrTrhSerGlnGln 188
Qy 451 -----ACCTTGCTCCAGCCGCCACCTACGCGGTGGCCATGATGCTTTGCCGCC 504
Db 189 ValGlyAlaIleAlaProAlaAlaGlyTyrGlyGlnGlyProAlaAlaValHisValAsp 208
Qy 505 TTGACCGATGCCAAGACTAGGAGCGATGATGATGATGGTCTGCTTTCTTCATG 564
Db 209 AlaTrhAlaIleTrhSerGlnGly-----AlaIleSerGlnGln 221
Qy 565 CAGGCTAGTATATACCAAGGGGATACACCGTTTGTCT 603
Db 222 GlyIseSerAlaAlaGlyGlyTyrGlyGlnGlnAla 234

RESULT 10
A55726
RNA-binding protein Ews - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A55726; S45007
R:Plongsatel, B.; Martel, M.G.; Thomas, G.; Delattre, O.
Article: Cloning and chromosome localization of the mouse Ews gene.
A:Reference number: A55726; MUID:95130099; PMID:7829090
A:Accession: A55726
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-655 <PDB>
A:Cross-references: GB:K79233; NID:g488512; PIDN:CAA55815.1; PID:g488513
A>Note: authors translated the codon TCA for residue 116 as Thr, GCC for residue 123 as
C:Genetics:
A:Gene: Ews
C:Superfamily: RNA-binding protein EWS; ribonucleoprotein repeat homology
C:Keywords: carcinogenesis; nucleus; RNA binding; tandem repeat
F:361-436/Domain: ribonucleoprotein repeat homology <RBM>

Alignment Scores:
Pred. No.: 1.54 Length: 655
Score: 98.50 Matches: 49
Percent Similarity: 34.88% Conservative: 11
Best Local Similarity: 28.49% Mismatches: 71
Query Match: 8.46% Indels: 41
DB: 1 Gaps: 9

US-09-809-545A-1_COPY_535_1143 (1-609) x A55726 (1-655)

Qy 85 TATGACGACGCGTGTGTCAGGCC---AACGAGAGGATCTTCATGTACACT 141
Db 29 TyrAlaGlnTrhTrh-----GlnAlaTyrGlyGlnGlnSerTyrGlyTyrTyrGly 45
Qy 142 GGGCCCATTCATCTGTATATCTTCGCAATGCTGCTTTCATATCCGGCCGCCACT 201
Db 46 GlnProTrhAspValSerTyrTrhGlnAlaGlnTrhTrhAlaTrhTyrGlyGlnTrhAla 65
Qy 202 GCTGACGTGATACGAGGGGCTCAGGCGCTGTCGAGCGCGTGTGCGACCGTATACAAACC 261
Db 66 TyrAlaTrhSerTyr-----GlyGlnProProTrhGlyTyrSerTrh 79
Qy 262 TTCAGAGCTGCGGGCCGCCCAACCCCAATCCCGGCTATGGCGGAGTGTGTATCAAG 321
Db 80 -----ProTrhAlaProGln-----AlaTyrSerGln 88
Qy 322 CCAGTGTATGCAATTAATGTGTACAGGGTGTACGCTGATACCGTACGCCAGCC 381
Db 89 ProValGlnGlyTyr-----GlyTrhGlyAlaTyrAspSerTrhTrhAla 103
Qy 382 ACCCTTGCACTGCTGCTGCTACAGTACAGTACGAGAGAGTTATGCTCCGAGCC 441
Db 104 TrhValTrhTrhTrhGlnAla-----SerTyrAlaAlaGlnSer 116


```

0Y      442 TACCAACAACAAGCTTGTCCAGCACCACCATTCAGGCGTTGGTGCCATGAATGCTTTTCG 501
Db      117 AAtyrglYThnGlPrcoAlaTyRProThryrGIgIn-----Gln 130
0Y      502 CCCTTGACCGATGCCAAGACTAGAGCACGTGATGATGTGGGTCTGCTCTTCTTCA 561
Db      131 ProThrAlaTrAlaProThrArGrProGlnAspGIyaNsLysPrcAlaGluThSerGln 150
0Y      562 TTGCAGCTAGTATATACCAAGGGGATACAAACCT 597
Db      151 ProGlnSer-----ThrGlyglYtyrAnGln 160

RESULT 11
T31328
fibroin - Chinese oak silkmoth
C:Species: Antherea pernyi (Chinese oak silkmoth)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T31328
R:Seturu, H., Tamura, T., Yukunhiro, K.
submitted to the EMBL Data Library, August 1998
A:Description: Characterization of the full length fibroin gene of a wild silkworm, Antea
A:Reference number: Z20995
A:Accession: T31328
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2639 <SEZ>
A:Cross-references: EMBL:AF083334; NID:g3450882; PID:g3450883; PIDN:AAC32606.1
C:Genetics:
I:introns: 14/3
```

Alignment Scores:	
Pred. No.:	1 45
Score:	98.50
Percent Similarity:	4.358
Best Local Similarity:	35.58
Query Match:	8.46%
DB:	2
Length:	2639
Matches:	37
Conservative:	6
Mismatches:	44
Indels:	18
Gaps:	4

US-09-809-545A-1_COPY_535_1143 (1-609) x T31328 (1-2639)

Oy	193	GCCCCACTGCTGCAGCTGCATATCCAGGGGGCTCACCTTCAGAGCCGGTGGACACCTG	252
Db	2532	AlaAlaAlaAlaAlaAlaAlaAlaAlaAlaGlySerArgArgSerGlyHisAspArg---Ala	2550
Oy	253	TACAAACCTTCAGAGCTGGCGGCCCAACCCCAATCCGACCTATGGCGGAGTACTG	312
Db	2551	TyrIleValaGlySerAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaGlyAlaGlyAlaSer	2570
Oy	313	TATCAAGCCAGCGATATGGCAATATA-----TTGCTACAGGGTGTAGCTGTCA	363
Db	2571	ArgIleValaGlyIleTyrGlyThrAspAspGlyPheValIleuAspLysGlyTyrAspSer	2590
Oy	364	TACCGTACGCCGCCACCCCTGCCCACTGCTGTGCTTCAAGTGAAGATTACGAGCA	423
Db	2591	GluGlySerAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaSerSerSerIleArg	2610
Oy	424	GT-----TTATGCTGGCAGCCCTACACACACA2ACTTGC	458
Db	2611	SerThrGluGlyHisProLeuLeuSerIleCysGlyArgPro-----Cys	2625
Oy	459	TCCAGCCCCCACC	470
Db	2626	SerHisSerHis	2629

```

RESULT 12
151920
mucin - rheus macaque (fragment)
C:Species: Macaca mulatta (rheus macaque)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000
C:Accession: 151920
R:Am. G.; Luo, G.; Wu, R.
Am. J. Respir. Cell Mol. Biol. 10, 546-551, 1994
#Title: Expression of MUC2 gene is down-regulated by vitamin A at the transcriptional level

```

A:Reference number: IS1920; MUID:94253322; PMID:8179918
A:Accession: IS1920
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 1-216 <RES>
A:Cross-references: EMBL:U00463; NID:g437054; PIDN:AAA20963.1; PID:g437055
C:Genetics:
A:Gene: MUC2

Alignment Scores:	
Pred. No.:	1 79
Score:	98 00
Percent Similarity:	35.88%
Best Local Similarity:	28.24%
Query Match:	8.42%
DB:	2
	9
Length:	216
Matches:	98
Conservative:	13
Mismatches:	60
Indels:	49
Gaps:	9

US-09-809-545A-1_COPY_535_1143 (1-609) X I51920 (1-216)

Oy	23	ACCCTACACCAATGCGTGGAAATTAATTCAGTTGTGGGGCGGCTACAGCCCGAC	82
Db	2	ThrProthPro----- lethrThrThrThrThrThrThrThrThrThrProthr	16
Oy	83	TCATAGCAGCAGCGTGTGTCGCCAGGCGCAACGAGAGATCTTCATGACAGT	142
Db	17	-----ProthrProthSerthrgIntnrProthr--Sert	28
Oy	143	GCCCCAGTTCATTGTATATATTCTTGCAATGCCGTGCTTCCATATCGGCGGCATG	202
Db	28	hrProlethrThrThrThrThrThralathrPro--ThrProthProthSerthrg	47
Oy	203	CTGCAGCTGCATACCCAGGCGCTACCTTGAGGCGGTGTCGA--CCGTGACACA	255
Db	47	IntnrProthr--SerthrProlethrThrThrThrThrThralathrProthrProthr	66
Oy	260	CCTTCAGAGCTGCGGCGCCCAACCCCAATCCGGGCTATGGCGAGTAGTATCAAG	318
Db	67	ProthrSerthrgIntnr-----ProthSerthrProlethr-----	79
Oy	320	AGCCAGTGTATGCAATTAATCTGACAGGGTGTATAGCTGATACCGTACGCCGAGC	375
Db	80	-----ThrThrThrThrThrThrThrThrThrThrThrThrProthrProthr	90
Oy	380	CCA-----CCCTCGCACTGCTGCTGCTCTACAGTACAGTACG	418
Db	91	ProthrSerthrgIntnrProthrProthProlethrThrThrThrThrThrThrValthr	110
Oy	419	GACAGTTATGCGCGGACCCCAACACACACAACTTGCTCCAGCCGCCACCP-----	472
Db	111	Pro-----ThrProthProthSerthrgIntnrProthrProthrProlethr	127
Oy	473	-----ACGGCGTGGTCCCA	487
Db	128	SerIntnrThrThralaMetThrPro	135

RESULT 13

hyperchemical:protein F41B6.11 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
 C:Accession: T31880
 R:Sammons, L.; Murray, J.
 submitted to the EMBL Data Library, July 1997
 A:Description: The sequence of C. elegans cosmid F41B6.
 A:Reference number: Z21095
 A:Accession: T31880
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-316 <SAM>
 A:Cross-references: EMBL:AF016448; PIDN:AA865959.1; GSPDB:GN00023; CESP:F41B6.11
 A:Experimental source: strain Bristol N2; clone F41B6
 C:Genetics:
 A:Gene: CESP:F41B6.11

A:Map position: 5
A:Introns: 9/2; 21/3
C:Superfamily: hydroxyproline-rich glycoprotein

Alignment Scores:

Pred. No.:	1.76	Length:	316
Score:	98.00	Matches:	46
Percent Similarity:	30.56%	Conservative:	9
Best Local Similarity:	25.56%	Mismatches:	65
Query Match:	8.42%	Indels:	60
DB:	2	Gaps:	7

US-09-809-545a-1_COPY_535_1143 (1-609) x T31880 (1-316)

```

QY 52 CCAGTTGTGGCGCGGTCTACAGCCCGCACTTCTATGACGAGCAGCGTGTGTGCGCAG 111
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 175 ProValProThrProValTyrGlnProThrTyr----- 185

QY 112 GCCAACGAGAGGATCTTCATGATGAGTGGCCCGCACTTCTATGATCTTCTGCA 171
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 186 ----- 187

QY 172 ATGCTGCTTTCATATCCGCGCGCACTGCTGACGCTGATACCGAGGGCTCACTT 231
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 188 ValProProCysProAlaProAlaAlaThrProValTyrAlaGlnProAlaPro----- 205

QY 232 CGAGCGCGTGTGCGACCGCTGTACAAACCTTCAGAGCTGCGCGCGC-----CCA 282
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 206 -----ArgProMetProValTyr-----AlaProAlaProAlaCysAlaGln 219

QY 283 CCCCCAATCCCGCGCTATGCGGAGTAGTGTATCAAGACCGAGTGTAGGCAATAATTG 342
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 220 PropCysProAlaTyrAsnProTyrGlyTyrGln-----GlyThrProGly 235

QY 343 CTACAGGCTGTACGCTGCATACCCCTACGCCAGCCACCCCTGCGCTGCT----- 399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 236 TyrAlaProAlaTyrAsnProTyrSerTyrValLeuArgSerProValAlaAlaAlaPro 255

QY 400 -----GCCCTACAGTGAC 411
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 256 GlyTyrAlaGlnGlyTyrAlaProAlaTyrAlaGlnGlyTyrAlaProAlaTyrAlaPro 275

QY 412 AGTACGAGCGATTATGCTGCCGACCCCTACACACACACTT-----GCTCCAGCCCC 468
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 276 SerTyrAlaProAlaTyrAlaProSerProPheArgAlaMetTyrLeuLeuGlySer 295

QY 469 ACCTACGGCTGTGCGCATGATGCTTTTGGCCCTTACCGATCGCAAGCATAGAGAC 528
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 296 AsnLysGlyLysLysAlaAlaAlaAlaAlaAlaAlaAlaAlaThrGlnGlyLysLysThr 315

```

RESULT 14

A26066 segmentation protein eve - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster
C:Date: 25-Oct-1987 #sequence_revision 25-Oct-1987 #text_change 24-Sep-1999

C:Accession: A26066; B26066; A26636
R:Macdonald, P.M.; Ingham, P.; Struhl, G.

Cell 47, 721-734, 1986
A:Title: Isolation, structure, and expression of even-skipped: a second pair-rule gene

A:Reference number: A26066; MUID:87051744; PMID:2877745

A:Accession: A26066
A:Molecule type: DNA

A:Residues: 1-376 <MAC>

A:Cross-references: GB:M14767; NID:9157386; PIDN:AAA285211; PID:9157387

A:Cross-references: GB:M14767
R:Fraser, M.; Hoey, T.; Rushlow, C.; Doyle, H.; Levine, M.

EMBO J. 6, 749-759, 1987
A:Title: Characterization and localization of the even-skipped protein of Drosophila.

A:Reference number: A26636; MUID:87218536; PMID:2884106

A:Molecule type: mRNA
A:Residues: 1-299, 'L', 301-376 <FRA>
A:Cross-references: GB:X05138; NID:97957; PIDN:CAA28784.1; PID:97958

C:Genetics: A:Gene: eve (even-skipped)

A:Cross-references: FlyBase:FBgn0000606
C:Superfamily: unassigned homeobox proteins; homeobox homology

C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:71-127/Domain: homeobox homology <HOX>

Alignment Scores:

Pred. No.:	1.74	Length:	376
Score:	98.00	Matches:	45
Percent Similarity:	28.65%	Conservative:	8
Best Local Similarity:	24.32%	Mismatches:	56
Query Match:	8.42%	Indels:	76
DB:	2	Gaps:	7

US-09-809-545a-1_COPY_535_1143 (1-609) x A26066 (1-376)

```

QY 64 GCGGTCTACAGCCCGCACTTCTATGACGAGCAGTGTGTGCGCAGCCAGCAGAG 123
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 138 AlaValTyrSerAspProAlaPheAlaAlaSerIleLeuGlnAlaAlaAlaSerVal 157

QY 124 GATCTTCCATGATACAGTGGCCCGCACTTCTATGATCTTCTGCAATGCTGCTTT 183
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 158 Gly-----MetProTyrPro 162

QY 184 CCATAT-----CGGCGCGCACTGCTGACGCTGCATACCGAGGGCTCACTTGAAGCGCT 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 163 ProTyrAlaProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 177

QY 241 GGTCCACCGTGTACACACCTTCAGAGCTGCGCGCCCAACC----- 285
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178 -----ValAlaThrAsnProMetLeuAlaThrGlyMetProMetGlyMetProGln 195

QY 286 -----CCATCCCGGCTATGCGGAGTAGTGTATCAAGAGCAGTATGAC 333
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 196 MetProThrMetGlyMetProGlyHisSerGlyHisAlaGlyHisProSerProTyrGly 215

QY 334 AATAAATTGCTACAGAGGTGTACGCTGCATACGCTACCGCTACGCCAGCCACCCCTGCACT 393
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 216 Gln-----TyrArgTyrThr----- 220

QY 394 GCTGCTCTACAGTACGATACGAGAGATTATGCTGCGACCCCTACACACACA 453
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221 -----ProTyrHisIlePro 225

QY 454 CTGCTCCAGCCCGCACTACGCGGCTGTGCTGCAATGCTTTGGCCCTTGACCAT 513
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 226 AlaArgProAlaProProHisProAlaGly----- 235

QY 514 GCCAAGACTAGAGCAATGCTGATGAGTGTGCTGCTTCTTCTTCTTACAGGCTACT 573
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 236 -----ProHisMetHisHisProHisMetMetGlySerSerAlaThrGlySer 251

QY 574 ATATACCAAGGGGA 588
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 252 SerTyrSerAlaGly 256

```

RESULT 15

T36581 probable transmembrane protein - Streptomyces coelicolor

C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T36581
R:Oliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, May 1999
A:Reference number: 221575

A:Accession: T36581

A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-811 <OLI>

A;Cross-references: EMBL:AL049826; PIDN:CAB42720.1; GSPDB:GN00070; SCOEDB:SCH24.16c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCH24.16c

Alignment Scores:
Pred. No.: 2.06 Length: 811
Score: 97.00 Matches: 49
Percent Similarity: 35.39% Conservative: 14
Best Local Similarity: 27.53% Mismatches: 63
Query Match: 8.33% Indels: 52
DB: 2 Gaps: 11

US-09-809-545a-1_copy_535_1143 (1-609) x T36581 (1-811)

```
QY 70 TACAGCCCCGCTTCTATGCGAGCGGCGTGTGCTGCAGCCACGAGAGGATCT 129
   :: ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 140 HisGluProAspAlaPheAlaHis-----LeuPheArgAspGlnGlnGlnGly 156
   |||
QY 130 TCC-----ATGTACAGTGGCCCGAGTTCACCTGTATATCTTGTCAATGCTGGCTTT 183
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 157 GlyHisProSerTyrAspAspGlnAlaAlaValProAlaProAlaProGly--- 175
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 184 CCATATCCGCGCCGCACTGCTGCAGCTGCATACCGAGGGGCTCACCTTCGAGGCCGTGCT 243
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 176 ProTyrGlnGlyAlaAlaGlyProGlyGlnTyr----- 186
   |||
QY 244 CGCACCGGTGTACAACAGCTTCAGAGCTGGGGCGCCGCCCAATCCGGGCTATGGC 303
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187 -----GlyAlaProProProProProProGlnTyrGly 198
   |||
QY 304 GGAGTAGTGTATCAAGAGCCAGTGTATGSCAATAATTGCTACAGAGGTGTTACGCTGCA 363
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 199 GlyHisGlnGlyProAspPro---TyrGlnAsnAlaGlnThrHisGlyGlnTyrGly 217
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 364 TACCGCTAAGCCAGCCAGCCCTGCGACGTGCTGCTACAGTGCAGTTCAGGACGA 423
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 218 -----SerAlaGlnTyrGlyGlySer----- 224
   |||
QY 424 GTTATGCTGCGGACCCCTTACACACACACTGTGCTCCAGCCCC---ACCTACGGCGTT 480
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 225 -----AlaAlaIleProGlnHisAla--AlaProAlaProGlnTyrAlaVal 241
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 481 GGTCGCAATG-----AATGCTTTTGGCGCCTTGACCGCATGCCAAGACTAGAGCCATGCT 534
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 242 AlaProThrProGlnAlaAlaGlnAlaProLeuGlnGlnProGlnProGlnProAlaAla 261
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 535 GATGATGTGGTCTGCTTCTTTCATTCAGGCTAGTATATATACCAAGGGGGA 588
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 262 Ser-----GlnAlaAlaProLysGlyGly 270
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
```

Search completed: March 12, 2003, 21:21:36
Job time : 29 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 12, 2003, 19:21:08 ; Search time 13.5 Seconds
(without alignments)
3742.086 Million cell updates/sec

Title: US-09-809-545a-1_COPY_535_1143

Perfect score: 1164
Sequence: 1 atgacataataaaagccgcttgcacat 609

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+n2p,model -DEV=xih
-Q=/cgn2_1/USPTO.spool/US09809545/rnat_11032003_130946_12306/app_query.fasta_1.775
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=n2p,rsp -MINMATCH=0.1 -LCOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blonsum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US09809545 @CGN 1.1 10 @rnat_11032003_130946_12306 -NCPU=6 -ICPU=3
-NO_XLXLY -NO_MMAP -HARBORRY -NBS_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FPGAPOP=6 -FPGAEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1020.5	87.7	396	1	A2BP_MOUSE
2	1009.5	86.7	397	1	A2BP_HUMAN
3	341.5	29.3	391	1	RBW9_HUMAN
4	116	10.0	3063	1	CA1C_HUMAN
5	115	9.9	3119	1	CA1C_MOUSE
6	103.5	8.9	1170	1	TSP2_BOVIN
7	103	8.8	3124	1	CA1C_CHICK
8	99.5	8.5	2003	1	NTC4_HUMAN
9	99	8.5	387	1	GAT6_CHICK
10	98.5	8.5	376	1	FXE1_HUMAN
11	98.5	8.5	655	1	EMS_MOUSE
12	98	8.4	376	1	HMEV_DROME
13	97	8.3	497	1	BRN1_RAT
14	97	8.3	1943	1	PC15_MOUSE
15	96.5	8.3	1964	1	NTC4_MOUSE
16	96	8.2	766	1	ERYA_DROME
17	95.5	8.2	1172	1	TSP2_HUMAN
18	94	8.1	386	1	HXAD_MOUSE

19	94	8.1	388	1	HXAD_HUMAN	P31271	homo sapien
20	93.5	8.0	320	1	CH36_CERCA	P17110	ceratilis c
21	93	8.0	197	1	MCS_MOUSE	P15265	mus musculus
22	93	8.0	495	1	BRN1_MOUSE	P31361	mus musculus
23	92.5	7.9	656	1	EMS_HUMAN	O01844	homo sapien
24	92	7.9	250	1	RECO_CAUCR	O98802	caulobacter
25	92	7.9	500	1	BRN1_HUMAN	P20264	homo sapien
26	91	7.8	755	1	RRE1_HUMAN	Q92766	homo sapien
27	90.5	7.8	450	1	AP2C_HUMAN	Q92719	homo sapien
28	90.5	7.8	1210	1	PER3_HUMAN	P56645	homo sapien
29	90.5	7.8	5179	1	MUC2_HUMAN	O02817	homo sapien
30	90	7.7	209	1	CCX1_HUMAN	O15255	homo sapien
31	89	7.6	463	1	Y102_MYCTU	O33951	mycobacteri
32	89	7.6	3149	1	TEGU_EBV	P03166	epstein-bar
33	88	7.6	401	1	HB9_HUMAN	P50219	homo sapien
34	88	7.6	474	1	VTP3_TVIV	P19275	thermoprote
35	88	7.6	585	1	PKNI_MYCTU	O10964	mycobacteri
36	88	7.5	830	1	SREC_HUMAN	O14162	homo sapien
37	87.5	7.5	224	1	Y091_NPVAC	P41479	autographa
38	87.5	7.5	362	1	NK2C_MOUSE	P97334	mus musculus
39	87.5	7.5	640	1	PPCM_HUMAN	Q16822	homo sapien
40	87	7.4	440	1	GAT4_RAT	P46152	rattus norv
41	87	7.5	465	1	FXD1_HUMAN	O16676	homo sapien
42	87	7.5	625	1	NIFA_AZOB	P30667	acospirillum
43	87	7.5	830	1	SREC_HUMAN	O14162	homo sapien
44	87	7.5	920	1	AD19_MOUSE	O35674	mus musculus
45	86.5	7.4	261	1	LEG3_RAT	P08699	rattus norv

ALIGNMENTS

RESULT 1

A2BP_MOUSE STANDARD; PRT; 396 AA.

AC Q9UJ43; Q9UJ45; 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ataxin-2-binding protein.
GN A2BP1 OR A2BP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Osada N., Kusuda Y., Tanuma R., Ito A., Hirata M., Sugano S.,
RA Hashimoto K.;
RT "Isolation of full-length cDNA clones from mouse brain cDNA library
made by oligo-capping method.";
RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 21-396 FROM N.A.
RX MEDLINE=20275619; PubMed=10814712;
RA Shihbata H., Huynh D.P., Pulst S.-M.;
RT "A novel protein with RNA-binding motifs interacts with ataxin-2.";
RL Hum. Mol. Genet. 9:1303-1313(2000).
CC -1- SUBUNIT: BINDS TO THE C-TERMINAL OF ATAXIN-2.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb.ch/announce/>
or send an email to license@isb.ch).
CC EMBL; AB041596; BAA95079.1;
CC EMBL; AF107204; AAF78292.1;
CC HSSP; P11940; IGV.
DR MGD; MGI:1926224; A2bp1.

```

DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; RRM; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS0102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
DR RNA-binding.
RM DOMAIN 116 192 RNA-BINDING (RRM).
FT CONFLICT 60 G->D (IN REF. 2).
FT CONFLICT 128 R->RFR (IN REF. 2).
FT CONFLICT 369 G->S (IN REF. 2).
SQ SEQUENCE 396 AA; 42590 MW; 9A0C5C95E8739F CRC64;

Alignment Scores:
Pred. No.: 3.48e-76 Length: 396
Score: 1020.50 Matches: 192
Percent Similarity: 96.06% Conservative: 3
Best Local Similarity: 94.58% Mismatches: 7
Query Match: 87.67% Indels: 1
DB: Gaps: 1

US-09-809-545A-1_COPY_535_1143 (1-609) x A2BP_MOUSE (1-396)

QY 1 ATGACTAATAAAGAGCCGTGAACCCCTACACCAATGCTGGAATTAATTCAGTTG 60
DB 195 MetThrAsnLysLysThrValAsnProTyrThrAsnGlyTyrLysLeuAsnProValVal 214
QY 61 GAGCGGCTACAGCCCGCACTTCTATGAGGACAGCGTCTGTGTGCGAGGCAACAG 120
DB 215 GlyAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuGlySerAlaAsnGln 234
QY 121 GAGGATCTTCATGATGACAGTGGCCCGCACTTGTATATACCTTCTCAATGCTGCG 180
DB 235 GluGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 254
QY 181 TTTCATATCCGCGCCGCACTGCTGACATACCGAGGGCTACCTTGAGGCGCT 240
DB 255 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 274
QY 241 GGTCCGACCGGTACAAACCTTCAGAGCTGGCGGCGCCCGCCCAATCCCGGCTAT 300
DB 275 GlyAlaGlnThrValTyrAsnThrPheArgAlaAlaAlaProProProAlaProAlaTyr 294
QY 301 GCGGAGTAGTGTATCAGAGCCAGTGTAGGCAATAAATGCTGCTCAAGGCTGTTACGCT 360
DB 295 GlyGlyValValTyrGlnAspGlyPheTyrGlyAlaAsp---IleTyrGlyGlyTyrAla 313
QY 361 GCATACCGCTTACGCCCGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
DB 314 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSerTyrGly 333
QY 421 CGAGTTTATGCTGCGAGCCCTACACCAACACCTGCTCCAGCCCGCCCACTGAGCGCTT 480
DB 334 ArgValTyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProAlaProTyrGlyVal 353
QY 481 GGTGCATGAATGCTTTTGGCGCTTACCGATGCGCAAGACTAGAGCAATGCTGATGAT 540
DB 354 GlyAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgGlyHisAlaAspAsp 373
QY 541 GTGGCTCTGCTTTCTTCTTCAATTCAGGCTAGTATATACCAAGGGGATACCAAGCTTTT 600
DB 374 ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrArgGlyGlyTyrAsnArgPhe 393
QY 601 GGTCCATAT 609
DB 394 AlaProTyr 396

RESULT 2
A2BP_HUMAN
ID A2BP_HUMAN STANDARD; PRT; 397 AA.
AC Q9NMB1; Q9NS20;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE Ataxin 2-binding protein.
GN A2BP1 OR A2BP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA WagaSuma M., Hosokawa T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuo Y.,
RA Nimomiya K., Iwayanagi T.;
RA "NEBO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
RN (2)
RP SEQUENCE OF 21-397 FROM N.A.
RX MEDLINE=20275619; PubMed=10814712;
RA Shibata H., Huynh D.P., Pulst S.-M.;
RT "A novel protein with RNA-binding motifs interacts with ataxin-2.";
RL Hum. Mol. Genet. 9:1303-1313(2000).
CC - SUBUNIT: BINDS TO THE C-TERMINAL OF ATAXIN-2.
CC - TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MUSCLE AND BRAIN.
CC - SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AK001027; BA91472.1; -
DR EMBL; AF107203; AAF78291.1; -
DR HSSP; P11940; 1CVJ.
DR MIM; 605104; -.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; RRM; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS0102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
RM RNA-binding.
FT DOMAIN 117 193 RNA-BINDING (RRM).
FT CONFLICT 92 92 A->T (IN REF. 2).
SQ SEQUENCE 397 AA; 42754 MW; E3E9060B868C79880 CRC64;

Alignment Scores:
Pred. No.: 2.77e-75 Length: 397
Score: 1009.50 Matches: 190
Percent Similarity: 95.57% Conservative: 4
Best Local Similarity: 93.60% Mismatches: 8
Query Match: 86.73% Indels: 1
DB: Gaps: 1

US-09-809-545A-1_COPY_535_1143 (1-609) x A2BP_HUMAN (1-397)

QY 1 ATGACTAATAAAGAGCCGTGAACCCCTACACCAATGCTGGAATTAATTCAGTTG 60
DB 196 MetThrAsnLysLysThrValAsnProTyrThrAsnGlyTyrLysLeuAsnProValVal 215
QY 61 GAGCGGCTACAGCCCGCACTTCTATGAGGACAGCGTCTGTGTGCGAGGCAACAG 120
DB 216 GlyAlaValTyrSerProLysPheTyrAlaGlyThrValLeuLeuGlySerAlaAsnGln 235
QY 121 GAGGATCTTCATGATGACAGTGGCCCGCACTTGTATATACCTTCTCAATGCTGCG 180
DB 236 GluGlySerSerMetTyrSerAlaProSerSerLeuValTyrThrSerAlaMetProGly 255
QY 181 TTTCATATCCGCGCCGCACTGCTGACATACCGAGGGCTACCTTGAGGCGCT 240

```

D	b	256	PheProTyrProAlaIalAthrAlaIalAalATyArgGlyAlaNIleuNhgGlyArg	275
O	y	241	GGTGGCACCCTGGTACAAACAACCTTCAGAGCTGGCGGCCCCCAACCCCAATCCGGGCAT	300
D	b	276	GlyAlgThrValTyArgSerThrPheArgAlaIalAalaProProProIleProAlaTyR	295
O	y	301	GGCCGAGNAGTGTAACGAAGCCAGAGTAGATGGCAATAATTGCTACAGGAGTTCACGCT	360
D	b	296	GlyGlyValValTyGlnAspGlyPheTyrglyAlaAsp--ileTyrglyGlyTyAla	314
O	y	361	GCATACCGCTACGCCACCCACCCACCCCTGCACCTGGCTGCCTACAGTACACGTTACGGA	420
D	b	315	AlaTyArgTyAlaGlnProThrProAlaThrAlaIalAalTySerAspSeryGly	334
O	y	421	CGAGTTATGCTGCGCAGCCCTTACCAACAACACTTGTCTCCAGCCCCCACTCGAGCGTT	480
D	b	335	ArgValTyAlaIalAaAspProTyTrhisHisAlaleuAlaProAlaproThTyrglyAla	354
O	y	481	GGTGCCATGATGCTTTTGGCCCTTGAACCCGATGGCAAGACTAGAGGACATGCTATAT	540
D	b	355	GlyAlaMetAlaAlaPhalaProLeuthrAlaPalaysthArgSerHisAlaAsp	374
O	y	541	GTGGCTTCGTTCTTCTTTCATTTGACGGCTAGTATATACCAAGGAGGATACACGTTTT	600
D	b	375	ValGlyLeuValLeuSerLeuGlnAlaSerIleTyArgGlyGlyTyArgSerArgPhe	394
O	y	601	GCTCCATAT 609	
D	b	395	AlaProTyR 397	
R	E	S	RESULT 3	
I	D	B	RM9 HUMAN STANDARD; PRT; 391 AA.	
A	C	O	043251; GUGM4;	
D	T	16-OC	T-2001 (Rel. 40, last created)	
D	T	16-OC	T-2001 (Rel. 40, last sequence update)	
D	T	15-JUN	-2002 (Rel. 41, last annotation update)	
D	E		Putative RNA-binding protein 9 (RNA binding motif protein 9).	
G	N		RM9.	
O	S		Homo sapiens (Human).	
O	C		Eukaryota; Metazoa; Chordata; Craniacea; Vertebrata; Euteleostomi;	
O	C		Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
O	X		NCBI_TaxId=9606;	
R	N	[1]	SEQUENCE FROM N.A.	
R	P		TISSUE=Placenta;	
R	A		Collins J.E., Burton J.;	
R	L		Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.	
R	P	[2]	SEQUENCE OF 86-391 FROM N.A.	
R	P		Whiteley M.;	
R	L		Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.	
R	L		-I- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).	
C	C		-----	
C	C		This SWISS-PROT entry is copyright. It is produced through a collaboration	
C	C		between the Swiss Institute of Bioinformatics and the EMBL outstation -	
C	C		the European Bioinformatics Institute. There are no restrictions on ite-	
C	C		use by non-profit institutions as long as its content is in no way com-	
C	C		modified and this statement is not removed. Usage by and for commercial	
C	C		entities requires a license agreement (see http://www.isb-sdb.ch/announce	
C	C		or send an email to license@sib-sdb.ch).	
C	C		-----	
D	R		EMBL; AL009266; CAA15842.1; ALT_SEQ.	
D	R		EMBL; AL049748; CAB63054.1; --	
D	R		HSSP; P11940; ICYU	
D	R		GeneW; HGNC:9906; RM9.	
D	R		InterPro; IPR00504; RNA_rec_mot.	
D	R		Pfam; PF00076; xrm; 1.	
D	R		SMART; SM00360; RRM; 1.	
D	R		PROSITE; PS0102; RRM; 1.	
D	R		PROSITE; PS00030; RRM_RNP_1; 1.	
K	M		RNA-binding.	
F	T		DOMAIN 112 188 RNA-BINDING (RRM).	

SO	SEQUENCE	391 AA;	42284 MW;	ED2B76C02D316A7B	CRC64;
<hr/>					
Alignment Scores:		1,466-20	Length:	391	
Pred. No.:	Score:	341.50	Matches:	93	
Percent Similarity:	48.66%	Conservative:	16		
Best Local Similarity:	41.52%	Mismatches:	63		
Query Match:	29.34%	Indels:	52		
DB:	1	Gaps:	10		
<hr/>					
US-09-809-545A-1_COPY_535_1143	(1-609)	x_REW9_HUMAN	(1-391)		
OY	1	ATGACGTAAATAAAAGCGCCGTGAACCCCTACACCAATAGCTGGAAATTAATTCAGTTGTG	60		
Db	191	MechTasmsysLysMetValThrProIleuAlaSerPheGlnAlaAspValSerLeu	21		
OY	61	GGCGCGGTACAGCCCCCACTTCTATGCAGGACCGGTGTGTGCCAGGCCAAC---	11		
Db	211	GIYAIAValIyrcIyProIleuIeuIyrAlaIaSerSerPheGlnAlaAspValSerLeu	23		
OY	118	---CAGAGGAGATTCTTCATGTACGTACGTGGCCCCCACTTACTT-----GTATTACTTCT	16		
Db	231	GIYAsnAspAlaAlaValProLeuSerIyArgIylGlyIleAsnThrTyrlleProleu	25		
OY	169	GCATGCTTGCTGCTTCCATATCCG---CGCGCACCTGCTCAGCTGATGATCCGAGGGGCT	22		
Db	251	IlelleProIlyPheProIyrrProthrIlaAlathrIraAlaAlaPheArgIyAla	27		
OY	226	CACCTTGAGAGCGGTGTGCACCGGTACAAACCTTCAGAGCTGGCGGCCCCACC	28		
Db	271	HisteuArgIyArgIyArgIythrValtyrIyAlaValArg--AlaValProProthr	28		
OY	286	CCAATCCCGGCTTAGCGCGGAGTACTGTATTAAGAAGCAAGTATGSCAATA-----	33		
Db	290	AlaIleProIatyrProIyAlIyAltyrGlnaspIyPheTyrcIy-AlaAspLeuty	30		
OY	338	-----AATGCTAAGG-----GTGTTACGCTGCA	36		
Db	309	rIleGluSerAlaAsnCysPheArgSerAsnArgValAspMetGlnProThrAspMetH	32		
OY	364	TACCGCTACGCCAGCCACCCCTCCACTGCTGCTGCT-----ACAGTGAC	41		
Db	329	SerIeuleuIeuGlnProGlnProProleuIeuGlnProleuGlnProleuThrAlah	34		
OY	412	AGTTACGAGCAAGTTATGCTCCGACCCCTTACCAACAACACTTGCTCCAGCCCCACC	47		
Db	349	rValMetAlaGlyCyethrGlnProThrProThr-----	36		
OY	472	TACGGCGTGTGTCACATGAATCTTTTGCGCCCTTGACCAATGCCAAGACTAGAGCCAT	53		
Db	361	-----MetProleuProleuPro--IeuAlaMetG	37		
OY	532	GCTGATGATGGGCTGCTGCTTCTTCTCATTCAGAGCTAGTATATCAAGGAGGATAC	59		
Db	370	ueuAlaIeuTrp-----ArgValTythrIouIuValalath	38		
OY	592	AACCGTTTTG 601			
Db	382	rAlaAspLeu 385			
<hr/>					
RESULT 4					
ID	CAIC_HUMAN	STANDARD,	PRT;	3063	AA.
AC	Q99715; Q99716;				
DT	15-JUL-1998 (Rel. 36, Created)				
DT	15-JUN-1998 (Rel. 36, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Collagen alpha I(XII) chain precursor.				
OS	COL12A1.				
GN	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
NCBI	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OK	NCBI_TaxID=9606;				

Oy	253	-----TACAAACCTTCAGACTGGGCG	276
Db	2117	LeuProPogLmaSnllleHslleSerAspGlutlrprYthrArghPheValSerTrp	2138
Oy	277	CCCCCACCACCCCAATCCCGGCCTTAGCGGAGTAGTGATGAAGAAGCACGTATGSGCAAT	336
Db	2137	AspProserProserProvalLeudcIlyTyrlvAlileValrlyrsProvalGIgysenShn	2150
Oy	337	AAATTGCTACAG-----GGTGGTTACGCTGCATACCGCTACGCCAACCCTCC	387
Db	2157	GluPromeetGuaIalapehVaIGlUmetEthSerYthrIreU-----	2171
Oy	388	GCCACTGCTGCTGCTACAGTGAAGACAGTACGACAGACAGTTATGCTGCCAACCCCTTACAC	447
Db	2171	-----	2171
Oy	448	CACAACACTTGCTGCAGCCCCCACCCTTACGCGGCTTGTCATGAATGCTTTGG-----	501
Db	2172	HlsAnleuAnProserThrThrTyAspAl-----AshValTyrlAlagInTyR	2188
Oy	502	-----CCCTTGACCATGACCAAGACTAGAGCCACTGTATGATGTG	543
Db	2189	AsperglYleuSerValProlerThraspGlnGLNgtYthrThr-----	2202
Oy	544	GGTCTCGTCTTCTTCTTCATTGACGAGCTATATACCAAGGGGAGATACACCGCTTT	600
Db	2203	--LeuTYrLeuAnValThrAspLeuLYsthrYclnIleIdLytpasPharPhe	2220
RESULT 5			
CALC_MOUSE			
ID	CALC_MOUSE	STANDARD;	PRT; 3119 AA.
AC	Q60847; P70322;		
DT	15-JUL-1998 (Rel. 36, Created)		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Collagen alpha 1(XII) chain precursor.		
GN	COL12A1.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS XI1A-1 AND		
RP	XI1B-1).		
RC	STRAIN=Swiss Webster, and C57BL/6J; TISSUE=Skin;		
RC	MEDLINE=96170761; Pubmed=8601035;		
RX	Boehme K., Li Y., Oh P.S., Olsen B.R.:		
RA	"Primary structure of the long and short splice variants cf mouse		
RT	collagen XII and their tissue-specific expression during embryonic		
RL	development.";		
RL	dev. Dyn. 204:432-445(1995).		
RN	[2]		
RP	PARTIAL SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS XI1A-2		
RP	AND XI1B-2).		
RC	STRAIN=C57BL/6J; TISSUE=Skin fibroblast;		
RC	MEDLINE=99348349; Pubmed=10419532;		
RA	Kania A.W., Reichenberger E., Baur S.T., KarimbuX N.Y., Taylor R.W.,		
RA	Olsen B.R., Nishimura I.:		
RT	"Structural variation of type XI collagen at its carboxyl-terminal		
RT	NCI domain generated by tissue-specific alternative splicing.";		
RL	J. Biol. Chem. 274:22053-22059(1999).		
CC	-1 FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-		
CC	CONTAINING FIBRILS. THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE		
CC	SURFACE OF THE FIBRILS. AND THE COL2 AND NC3 DOMAINS MAY BE		
CC	LOCALIZED IN THE PERIFIBRILLAR MATRIX (BY SIMILARITY).		
CC	-1 SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 kDa OF		
CC	NONREPEAT-HELICAL SEQUENCES (BY SIMILARITY).		
CC	-1 ALTERNATIVE PRODUCTS: 4 ISOFORMS; XI1A-1 (SHOWN HERE), XI1A-		
CC	2/ER#K, XI1B-1 AND XI1B-2; ARE PRODUCED BY ALTERNATIVE SPLICING.		
CC	THE FINAL TISSUE FORM OF COLLAGEN XII MAY CONTAIN HOMOTRIMERS OR		
CC	ANY COMBINATION OF THE VARIOUS ISOFORMS.		
CC	-1 TISSUE SPECIFICITY: HIGHEST EXPRESSION IN TENDONS, PERICHONDRUM,		
CC	SKIN, CORNEA, SCLERA, BLOOD VESSELS, AND PERITONEUM.		

	-I-	DEVELOPMENTAL STAGE: THE LONG NC3 XIIA ISOFORMS ARE PREDOMINANT AT EARLY STAGES (ED7 AND I1), AT LATER STAGES OF DEVELOPMENT (ED15 AND I7). THE SHORT NC3 XIIB FORM BECOMES THE MAJOR FORM. AS THE SHORT NC3 FORMS BECOME THE MAJOR PRODUCT, THE LONG SPICE VARIANT CONTINUES TO BE EXPRESSED IN SEVERAL TISSUES, EVEN AFTER BIRTH.
	CC	THE LONG NC1 ISOFORMS, XIIIA-1 AND XIIB-1, PEAK IN 15-DAY OLD EMBRYOS AND DECREASE IN 17-DAY OLD ONES. THE EXPRESSION OF THE SHORT NC1 FORM XIIB-2 REMAINS CONSTANT THROUGHOUT LATE STAGES OF EMBRYONIC DEVELOPMENT (ED15 AND ED17).
	CC	-I- PMM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT EACH END (BY SIMILARITY).
	CC	-I- PMM: PROLINS ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY SIMILARITY).
	CC	-I- PMT: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY). ONLY ISOFORM XIIA-2 IS A PROTOGLUCAN.
	CC	-I- SIMILIARTY: BELONGS TO THE FIBRILL-ASSOCIATED COLLAGENS WITH INTERRUPTED HELICES (PACIF) FAMILY.
	CC	-I- SIMILIARTY: CONTAINS 4 VWFA DOMAIN.
	CC	-I- SIMILIARTY: CONTAINS 18 FIBRONECTIN TYPE III-LIKE DOMAINS.
	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outpost - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (http://www.isb-ch/announce/or_send_an_email_to_license@isb-sib.ch). -----
DR	EMBL; U25652;	AAA99719.1; ALT_SEQ.
DR	EMBL; U57095;	AAB07047.1; --
DR	HSSP; P02751;	IENA.
DR	MCD; MGI:	B84448; Col12a1.
DR	InterPro; IPRO00087;	Collagen.
DR	InterPro; IPRO03961;	FN III.
DR	InterPro; IPRO03962;	FNIII_repeat.
DR	InterPro; IPRO03129;	TSPN.
DR	Pfam; PF00041;	fni3; 18.
DR	Pfam; PF00092;	vwa; 4.
DR	Pfam; PF01391;	Collagen; 4.
DR	pfdm; pf02210;	tspn; 1
DR	PRINTS; PR00014;	ENTYPE.III.
DR	PRINTS; PR00453;	VWFDOMAIN.
DR	SMART; SMOO060;	FN3; 16.
DR	SMART; SMO0210;	TSPN; 1.
DR	SMART; SMO0327;	VWA; 4.
DR	PROSITE; PS50234;	VWFA; 4.
KW	Extracellular matrix;	Connective tissue; Repeat; Hydroxylation; Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.
FT	SIGNAL	1..24
FT	CHAIN	25..3119
FT	DOMAIN	25..114
FT	DOMAIN	140..316
FT	DOMAIN	332..423
FT	DOMAIN	444..620
FT	DOMAIN	634..725
FT	DOMAIN	726..816
FT	DOMAIN	817..907
FT	DOMAIN	908..1002
FT	DOMAIN	1003..1089
FT	DOMAIN	1090..1182
FT	DOMAIN	1203..1375
FT	DOMAIN	1388..1474
FT	DOMAIN	1480..1568
FT	DOMAIN	1569..1652
FT	DOMAIN	1659..1757
FT	DOMAIN	1758..1848
FT	DOMAIN	1849..1938
FT	DOMAIN	1939..2029
FT	DOMAIN	2030..2120
FT	DOMAIN	2121..2208
FT	DOMAIN	2209..2297
		COLLAGEN ALPHA 1(XII) CHAIN. VWFA 1.. VFRA 2.. FIBONECTIN TYPE-I II 3.. FIBONECTIN TYPE-II II 13.. FIBONECTIN TYPE-III 5.. FIBONECTIN TYPE-III 6.. FIBONECTIN TYPE-III 7.. FIBONECTIN TYPE-III 8.. VFMA 3.. FIBONECTIN TYPE-III 9.. FIBONECTIN TYPE-III 10.. FIBONECTIN TYPE-III 11.. FIBONECTIN TYPE-III 12.. FIBONECTIN TYPE-III 13.. FIBONECTIN TYPE-III 14.. FIBONECTIN TYPE-III 15.. FIBONECTIN TYPE-III 16.. FIBONECTIN TYPE-III 17.. FIBONECTIN TYPE-III 18..

FT	DOMAIN	2329	2501	VWFA 4.
FT	DOMAIN	2456	2751	NONHELICAL REGION (NC3).
FT	DOMAIN	2752	2899	TRIPLE-HELICAL REGION (COL2)
FT	DOMAIN			WITH 1 IMPERFECTION.
FT	DOMAIN	2900	2942	NONHELICAL REGION (NC2).
FT	DOMAIN	2943	3045	TRIPLE-HELICAL REGION (COL1)
FT	DOMAIN			WITH 2 IMPERFECTIONS.
FT	BINDING	3046	3119	NONHELICAL REGION (NC1).
FT	BINDING	802	802	TO CHONDROITIN SULFATE (POTENTIAL).
FT	BINDING	893	893	TO CHONDROITIN SULFATE (POTENTIAL).
FT	BINDING	985	985	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	866	868	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	2784	2786	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	2896	2898	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	2945	2945	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	2948	2948	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	2951	2951	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	2960	2960	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	2966	2966	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	2969	2969	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	2972	2972	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	2984	2984	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	3001	3001	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	3004	3004	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	3015	3015	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	3024	3024	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	3027	3027	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	3030	3030	HYDROXYLATION (BY SIMILARITY).
FT	DOMAIN	869	872	POLY-THR.
FT	CARBOHYD	704	704	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1769	1769	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2212	2212	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2533	2533	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2684	2684	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPPLIC	25	1190	MISSING (IN ISOFORM XIIB-1 AND ISOFORM XIIB-2).
FT	VARSPPLIC	3062	3064	Epy -> GSG (IN ISOFORM XIIA-2 AND ISOFORM XIIB-2).
FT	VARSPPLIC	3065	3119	MISSING (IN ISOFORM XIIB-2).
FT	SEQUENCE	3119 AA;	340239 MM;	9B1F99C8A6A3251 CRC64;

Alignment Scores:

Pred. No.: 0.0567

Score: 115.00

Percent Similarity: 33.52%

Best Local Similarity: 24.43%

Query Match: 9.88%

OB: 1

Length: 3119

Matches: 43

Conservative: 16

Mismatches: 59

Indels: 58

Gaps: 7

[illegible]

```

Db 2177 ----- 2177
QY 448 CACACACTTGCTCCAGCCCCCAGCTAGGGCGTTGGCCATGATGCTTT----- 498
Db 2178 HisenLeuansProSerThrThrTyraSpalSerValTyraIaGlnTyraSpSerGly 2197
QY 499 -----GCGCCCTTGACCGATGCGCAACACTYAGAGACCATGCTGATGATGGTCTCGTT 552
Db 2198 LeuSerValProLeuThrAspGlnGlyThrThr-----LeuTyr 2210
QY 553 CTTCCTTCATTCAGGAGGCTAGTATACCAAGGGGATACCAACGCTTT 600
Db 2211 LeuansValThrAspLeuTyrThrTyGlnValGlyTyraSpThrPhe 2226

RESULT 6
TSP2_BOVIN STANDARD; PRT; 1170 AA.
ID TSP2_BOVIN
AC Q95116; Q28180;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thrombospondin 2 precursor (corticotropic-in-induced secreted protein)
GN (CISP).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Danik M., Chinn A., Lafeuillade M., Keramidas M., Aguesse-Germon S.,
RA Pehost A., Chen H., Mosher D., Chambaz E.M., Feige J.J.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-522 FROM N.A.
RX MEDLINE=96331130; PubMed=8698834;
RA Lafeuillade B., Pellerin S., Keramidas M., Danik M., Chambaz E.M.,
RA Feige J.J.;
RT "Opposite regulation of thrombospondin-1 and corticotropin-induced
RT secreted protein/thrombospondin-2 expression by adrenocorticotrophic
RT hormone in adrenocortical cells.";
RT J. Cell. Physiol. 167:164-172(1996).
RN [3]
RP SEQUENCE OF 318-831 FROM N.A.
RC TISSUE=Aortic endothelium;
RA Zafar R.S., Moll Y.D., Womack J.F., Walz D.A.;
RT "Cloning and sequencing of bovine thrombospondin stimulatory effect of
RT TGF-beta.";
RT Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
RL -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
RL CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
RL LAMININ AND TYPE V COLLAGEN.
CC -1- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 WPC DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -1- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as their content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X96540; CAA65385.1; -
DR EMBL; X87620; CAA60952.1; -
DR HSSB; P00740; IEDM.
DR InterPro; IPR000561; EGF-like.

```

DR	InterPro:	IPR001881;	EGF Ca.
DR	InterPro:	IPR001791;	Laminin_G.
DR	InterPro:	IPR000864;	TSPN.
DR	InterPro:	IPR003128;	TSPN.
DR	InterPro:	IPR001007;	VWF_C.
DR	InterPro:	IPR003367;	tsp_3.
DR	Pfam:	PF00008;	EGF_2.
DR	Pfam:	PF00090;	tsp_1; 3.
DR	Pfam:	PF00093;	vwc_1.
DR	Pfam:	PF02210;	TSPN_1.
DR	Pfam:	PF02412;	tsp_3; 9.
DR	SMART:	SM00181;	EGF_3.
DR	SMART:	SM00209;	TSP1; 3.
DR	SMART:	SM00210;	TSPN; 1.
DR	SMART:	SM00214;	VWC; 1.
DR	PROSITE:	PS00022;	EGF_1;
DR	PROSITE:	PS01186;	EGF_2; 1.
DR	PROSITE:	PS50092;	TSP1; 3.
KM	Glycoprotein:	Cell adhesion;	Calcium-binding; Heparin-binding; Repeat;
KM	EGF-like domain:	Signal.	
FT	SIGNAL	1	18
FT	CHAIN	19	1170
FT	DOMAIN	19	232
FT	DOMAIN	318	375
FT	DOMAIN	379	430
FT	DOMAIN	435	491
FT	DOMAIN	492	546
FT	DOMAIN	547	587
FT	DOMAIN	588	645
FT	DOMAIN	646	690
FT	DOMAIN	723	758
FT	DOMAIN	759	781
FT	DOMAIN	793	817
FT	DOMAIN	818	840
FT	DOMAIN	841	878
FT	DOMAIN	879	914
FT	DOMAIN	915	950
FT	DOMAIN	951	1170
FT	SITE	926	928
FT	DISULFID	266	266
FT	DISULFID	270	270
FT	DISULFID	551	562
FT	DISULFID	556	572
FT	DISULFID	575	586
FT	DISULFID	592	608
FT	DISULFID	599	617
FT	DISULFID	620	644
FT	DISULFID	650	663
FT	DISULFID	657	676
FT	DISULFID	678	689
FT	CARBOHYD	151	151
FT	CARBOHYD	316	316
FT	CARBOHYD	330	330
FT	CARBOHYD	435	435
FT	CARBOHYD	582	582
FT	CARBOHYD	708	708
FT	CARBOHYD	936	936
FT	CARBOHYD	1067	1067
FT	CONFLICT	535	535
FT	CONFLICT	748	748
EQ	SEQUENCE	1170 AA;	129862 MW; 9GCFIBF55B89A051 CRC64;

Alignment Scores:	
Pred. No.:	0.479
Score:	103.50
Percent Similarity:	34.12%
Best Local Similarity:	24.17%
Query Match:	8.69%
DB:	1
	Gaps:
	12
	Length:
	1170
	Matches:
	51
	Conservative:
	21
	Mismatches:
	69
	Indels:
	71
	Gaps:
	12

US-09-809-545A-1_COPY_535_1143 (1-609) X TSP2_BOVIN (1-1170)

```

QY      12  AAGGCGGTGAACCCCTTACACACATGGCTGTGAATTAAATCAAGCTGTGGGCGCGGCTTA 71
           ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      380  GluGlyTrpSerProTrpAlaGluTrpThrGluCysSerAlaThrCysGlySerGlyThr 359
QY      72  CAGCCCGGACT-----TGTATGAC 91
           |||||
Db      400  GlnGlnArgGlyArgSerCysAspValThrSerAsnThrCysLeuGlyProSerIleGln 419
QY      92  GCACGGTGGTGT-----TGTGCCAGGCCCAACGAGAGGAT----- 127
           |||||
Db      420  ThrArgAlaCysSerLeuGlyArgCysAspHisArgIleArgGlnAspGlyGlyTrpSer 439
QY      128  -----CTTCCAACTGTACATGGGCGCCAGTTCATCTGTAT 160
           ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      440  HistTrpSerProTrpSerSerCysSerValThrCysGlyValGlyAsnVal-ThrArgIle 459
QY      161  ATACTTGTCAATGCGCTGGCTT--TCCATATCCGGCGCGGCGACTGTGCAGCTGCATACC 217
           |||||::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      459  eArgLeuCyAsnSerProValProGlnMetGlyIlyArgSerCys----- 474
QY      218  GAGGGGCTCACCTTTCAGAGCGCGGTGTGCACCGTATCAACACCTTCAAGAGCTGC--GG 274
           ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      475  -LysGlySerGly-----ArgGluThrLysAlaCysGlnGly 486
QY      275  CGCCCCCAACCCCAATCCCGGCGCTATGGCGGAGTACAGTATCAAGAGCCGCTGATGGCA 334
           |||||::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      486  YProProCysProValAlaAspGlyArgTrp-----SerProTrp-- 498
QY      335  ATAAATTGCTACAGGAGTGTTCAGCTGCATACCGCTACGCCCGACCCCGCTGCACACTG 394
           ::::|::|::|
Db      499  -----SerProTrpSerAlaCysThrVal-----Th 507
QY      395  CTGCTGCCTACAGTGCACGTTACGAGACGAGTATTATGCTGCCAGCCCTACACACAC-- 452
           |||||::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      507  rCyAlaGlyIlyIleArgGluArgThrArgValaCysAsnSerProGluProGlnHisGly 527
QY      453  -----ACTTGCTCCAGAGCCCGCCACCTACGCGCGTGTGGTCGACATGAAGCTTTGGCC 502
           ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      527  yGlyLysAspCysValaGlyIlyAlaLysGluGlnGlnMetCysAsnArgLysSerCysPr 547
QY      503  CCTTGACCGATGCCAAGACTACGAGCCCATGC 533
           ::::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      547  oIleAspGlyCysLeuSer---AsnProCys 556

RESULT 7
CALC_CHICK
ID      CAIC_CHICK      STANDARD;      PRT; 3124 AA.
AC      P13944; 004509;
DT      01-JAN-1990 (Rel. 13, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Collagen alpha 1(XII) chain precursor (Fibrochimerin).
GN      COL12A1.
OS      Gallus gallus (Chicken).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC      Gallus.
OX      NCBI_TaxID=9031;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=White Leghorn;
RX      MEDLINE=92011862; PubMed=1918137;
RA      Yanagita M., Yamada K.M., Yamada S.S., Shinomura T., Tanaka H.,
RA      Nishida Y., Obata M., Kimata K.;
RT      "The complete primary structure of type XII collagen shows a chimeric
RT      molecule with reiterated fibronectin type III motifs, von Willebrand
RT      factor IX motifs, a domain homologous to a noncollagenous region of
RT      type IX collagen, and short collagenous domains with an Arg-Gly-Asp
RT      site."
RT      J. Cell Biol. 115:209-221(1991).
RP      [2]
RP      SEQUENCE OF 2456-3124 FROM N.A., AND SEQUENCE OF 2772-2794 AND
RP      2846-2873.

```

RP SEQUENCE OF 2456-3124 FROM N.A., AND SEQUENCE OF 2772-2794 AND
RP 2846-2873.

RP 2846-2873

RX MEDLINE=90062079; PubMed=2584192;
 RA Gordon M.K., Gerecke D.R., Dublet B., van der Rest M., Olsen B.R.;
 RT "Type XII collagen. A large multidomain molecule with partial
 RT homology to type IX collagen."
 RL J. Biol. Chem. 264:19772-19778(1989).
 RN [3]
 RP SEQUENCE OF 2960-3076 FROM N.A.
 RX MEDLINE=87317590; PubMed=3476925;
 RA Gordon M.K., Gerecke D.R., Olsen B.R.;
 RT "Type XII collagen: distinct extracellular matrix component
 RT discovered by cDNA cloning."
 RL Proc. Natl. Acad. Sci. U.S.A. 84:6040-6044(1987).
 RN [4]
 RP SEQUENCE OF 1-1283 FROM N.A. (SHORT FORM), AND ALTERNATIVE SPLICING.
 RC TISSUE-EMBRYO;
 RX MEDLINE=93042014; PubMed=1420368;
 RA Trueb J., Trueb B.;
 RL Blochim. Biophys. Acta 1171:97-98(1992).
 RN [5]
 RP ALTERNATIVE SPLICING.
 RX MEDLINE=95370352; PubMed=7642694;
 RA Koch M., Bohmann B., Mathiesen M., Hagios C., Trueb B., Chiquet M.;
 RT "Large and small splice variants of collagen XII: differential
 RT expression and ligand binding."
 RL J. Cell Biol. 130:1005-1014(1995).
 CC -1- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-
 CC CONTAINING FIBRILS. THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE
 CC SURFACE OF THE FIBRILS, AND THE COL2 AND NC3 DOMAINS MAY BE
 CC LOCALIZED IN THE PERIFIBRILLAR MATRIX.
 CC -1- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 KDa OF
 CC NONTRIPLE-HELICAL SEQUENCES.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE FINAL TISSUE
 CC FORM OF COLLAGEN XII MAY CONTAIN HOMOTRIMERS OF EITHER THE LONGER
 CC OR THE SHORTER ISOFORM OR ANY COMBINATION OF LONG AND SHORT
 CC ISOFORM CHAINS. ONLY THE LONG VARIANT IS A PROTEOLYCAN. THE LARGE
 CC ISOFORM HAS MORE RESTRICTED EXPRESSION IN EMBRYONIC TISSUE THAN
 CC THE SMALL.
 CC -1- TISSUE SPECIFICITY: TYPE XII COLLAGEN IS PRESENT IN TENDONS,
 CC LIGAMENTS, PERICHONDRUM, AND PERIOSTEUM, ALL DENSE CONNECTIVE
 CC TISSUES CONTAINING TYPE I COLLAGEN.
 CC -1- DOMAIN: THIS SEQUENCE DEFINES FIVE DISTINCT DOMAINS, TWO TRIPLE-
 CC HELICAL DOMAINS (COL1 AND COL2) AND THREE NONTRIPLE-HELICAL
 CC DOMAINS (NC1, NC2, AND NC3).
 CC -1- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT
 CC EACH END.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
 CC INTERRUPTED HELICES (FACIT) FAMILY.
 CC -1- SIMILARITY: CONTAINS 4 VMFA DOMAINS.
 CC -1- SIMILARITY: CONTAINS 18 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL, D00824; BAA00701.1; -
 DR EMBL, X61024; CAA43358.1; -
 DR EMBL, M17375; AAA48718.1; -
 DR EMBL, J05137; AAA48635.1; -
 DR EMBL, X67327; CAA47744.1; -
 DR PIR, A28037; A28037.
 DR PIR, A34485; A34485.
 DR HSSP, P20701; LIFA.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR003961; FN_III.

DR InterPro; IPR003962; FNIII_repeat.
 DR InterPro; IPR003129; TSPN.
 DR InterPro; IPR002035; VMF_A.
 DR Pfam; PF00041; fn3; 17.
 DR Pfam; PF00092; vma; 4.
 DR Pfam; PF01391; collagen; 4.
 DR Pfam; PF02210; TSPN; 1.
 DR PRINTS; PR00014; FNTYPEIII.
 DR PRINTS; PR00453; VMFADOMAIN.
 DR SMART; SM00060; FN3; 17.
 DR SMART; SM00210; TSPV; 1.
 DR SMART; SM0327; VMA; 4.
 DR PROSITE; PS50234; VMFA; 4.
 KM Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.
 FT SIGNAL 1 24
 FT CHAIN 25 3124
 FT DOMAIN 25 114
 FT DOMAIN 139 311
 FT DOMAIN 332 425
 FT DOMAIN 439 615
 FT DOMAIN 629 720
 FT DOMAIN 721 811
 FT DOMAIN 812 904
 FT DOMAIN 905 998
 FT DOMAIN 999 1085
 FT DOMAIN 1086 1178
 FT DOMAIN 1199 1371
 FT DOMAIN 1384 1473
 FT DOMAIN 1474 1565
 FT DOMAIN 1566 1654
 FT DOMAIN 1655 1755
 FT DOMAIN 1756 1846
 FT DOMAIN 1847 1936
 FT DOMAIN 1937 2027
 FT DOMAIN 2028 2118
 FT DOMAIN 2119 2206
 FT DOMAIN 2207 2295
 FT DOMAIN 2327 2500
 FT DOMAIN 2455 2750
 FT DOMAIN 2751 2902
 FT DOMAIN 2903 2945
 FT DOMAIN 2946 3048
 FT DOMAIN 3049 3124
 FT DOMAIN 3086 3096
 FT DOMAIN 3111 3123
 FT BINDING 797 797
 FT BINDING 890 890
 FT BINDING 981 981
 FT SITE 2438 2440
 FT SITE 2899 2901
 FT CAROXYD 32 32
 FT CAROXYD 1006 1006
 FT CAROXYD 1032 1032
 FT CAROXYD 1044 1044
 FT CAROXYD 1512 1512
 FT CAROXYD 1767 1767
 FT CAROXYD 2210 2210
 FT CAROXYD 2273 2273
 FT CAROXYD 2532 2532
 FT CAROXYD 2683 2683
 FT VARSPIC 25 1188
 FT CONFLICT 1258 1258
 FT CONFLICT 1264 1264
 FT CONFLICT 2759 2759
 FT CONFLICT 2803 2803
 FT CONFLICT 2977 2977
 FT CONFLICT 3075 3076
 SQ SEQUENCE 3124 AA; 340578 MW; 094285AFB7F346CF CRC64;

Alignment Scores:

DR HSPP; P08709; 18F9.
 DR Genew; HGNC:7884; NOTCH4.
 DR MIM; 164951; .
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Anx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000742; EGF-2.
 DR InterPro; IPR01881; EGF_Ca.
 DR InterPro; IPR01438; EGF-1.
 DR InterPro; IPR00800; Notch.
 DR Pfam; PF00008; EGF; 26.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00066; notch; 2.
 DR PRINTS; PRO0010; EGFBLD.
 DR PRINTS; PRO0011; EGFBLDIN.
 DR PRINTS; PRO0012; FNTYPE.
 DR SMART; SM00248; ANK; 5.
 DR SMART; SM00179; EGF_Ca; 11.
 DR SMART; SM00001; NL; 2.
 DR SMART; SM00004; NL; 2.
 DR PROSITE; PS50088; ANK_REPEAT; 5.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 11.
 DR PROSITE; PS00022; EGF-1; 28.
 DR PROSITE; PS01186; EGF-2; 21.
 DR PROSITE; PS01187; EGF_Ca; 9.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
 KW Tissue; Platelet expansion; Alternative splicing.
 FT SIGNAL 1 23
 FT CHAIN 24 2003
 FT CHAIN 1432 2003
 FT CHAIN 1467 2003
 FT DOMAIN 24 1447
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 118 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 274
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 806 842
 FT DOMAIN 844 880
 FT DOMAIN 882 928
 FT DOMAIN 930 966
 FT DOMAIN 968 1004
 FT DOMAIN 1006 1044
 FT DOMAIN 1046 1085
 FT DOMAIN 1087 1126
 FT DOMAIN 1130 1171
 FT DOMAIN 1472 1476
 FT REPEAT 1165 1212
 FT REPEAT 1213 1246
 FT REPEAT 1247 1286
 FT REPEAT 1633 1665

FT REPEAT 1666 1698 ANK 2.
 FT REPEAT 1700 1732 ANK 3.
 FT REPEAT 1733 1765 ANK 4.
 FT REPEAT 1766 1798 ANK 5.
 FT DISULFID 28 41 BY SIMILARITY.
 FT DISULFID 35 51 BY SIMILARITY.
 FT DISULFID 53 62 BY SIMILARITY.
 FT DISULFID 68 80 BY SIMILARITY.
 FT DISULFID 74 103 BY SIMILARITY.
 FT DISULFID 105 114 BY SIMILARITY.
 FT DISULFID 122 133 BY SIMILARITY.
 FT DISULFID 127 143 BY SIMILARITY.
 FT DISULFID 145 154 BY SIMILARITY.
 FT DISULFID 160 171 BY SIMILARITY.
 FT DISULFID 165 180 BY SIMILARITY.
 FT DISULFID 182 191 BY SIMILARITY.
 FT DISULFID 198 211 BY SIMILARITY.
 FT DISULFID 205 220 BY SIMILARITY.
 FT DISULFID 222 231 BY SIMILARITY.
 FT DISULFID 238 249 BY SIMILARITY.
 FT DISULFID 243 262 BY SIMILARITY.
 FT DISULFID 264 273 BY SIMILARITY.
 FT DISULFID 280 291 BY SIMILARITY.
 FT DISULFID 285 300 BY SIMILARITY.
 FT DISULFID 302 311 BY SIMILARITY.
 FT DISULFID 318 332 BY SIMILARITY.
 FT DISULFID 326 341 BY SIMILARITY.
 FT DISULFID 343 352 BY SIMILARITY.

Alignment Scores:
 Pred. No.: 1.04 Length: 2003
 Score: 99.50 Matches: 51
 Percent Similarity: 30.77% Conservative: 17
 Best Local Similarity: 23.08% Mismatches: 62
 Query Match: 8.55% Indels: 91
 DB: Gaps: 14

US-09-809-545A-1_COPY_535_1143 (1-609) x NTc4_HUMAN (1-2003)

QY 57 TGTGGCGCGGCTCTACAGCCCGCACTTCTA----- 86
 DB 918 CysProProGlyPheGlnGlySerIleuGlyGlnPheHisValaenProGlySerGluSerArg 937
 QY 87 --TGCAGGACGAGTGTCTGTGTGTCAGAGCCCAAGAGGAGATTTCCATGTACAGTGG 143
 DB 938 ProGlyGlnaenGlyAlaThrCysmetaGlnProSerGly----- 951
 QY 144 CCCGAGTTCACCTGTATATCTTCGCAATGCGCTTCCATATCGGCGCCGCACTGC 203
 DB 952 -----TyleuGlyGlnCysAlaProGlyTyraSpGlyGlnaenCys 965
 QY 204 TGCAGCTGCATACGAGGAGGCTCACCTTCGAGGCGGTGTCGACCGTGTACACACCTT 263
 DB 966 Ser-----LysGluLeu 969
 QY 264 CAGAGCTGC-----GGCGCCCCCACCCTCAATCCGGCCTA 299
 DB 970 AspAlaCysGlnSerGlnProCysHisaenHisGlyThrCysThProLysProGlyGly 989
 QY 300 TGGCGGAGTAGTGATTCAGAGCCGAGTATGCGCAATAATGCTACAGGGGTGTACGC 359
 DB 990 PheHisCysAlaCysProProGlyPheVal-----GlyLeuArg 1002
 QY 360 TGC-----ATACCGCTACGCGCCAGCCACCC----- 386
 DB 1003 CysGluGluGlyAspValaenGlyCysLeuAspGlnProCysHisProThrGlyThrAlaAla 1022
 QY 387 TGCACCTGCTGCTGCTGCTACAGTACGACGAGTATAGCTGCCGACCCCTTACCA 446
 DB 1023 CysHis-----SerLeuAlaenAlaPheTyCysGlnCysLeuPro 1036
 QY 447 CCACACACTTGCTCCAGCCCGCCACCTAGGCGCTTGCTGTCATGAATGCTTTGGCCCTT 506
 DB 1036 CysHis-----SerLeuAlaenAlaPheTyCysGlnCysLeuPro 1036


```

Db 1037 GlyHisThr-----GlyGlnTrpCysGlu-----ValGluIle 1047
QY 507 GACCGATGCCAAGACTAGACCATGC-----TGATGATGTGG 545
Db 1048 AsproCysHisSer---GlnProCysPheHisGlyThrCysGluAlaThrAlaGly 1066
QY 546 TCTGCTTTCTTTCATTCATGACGGCTAGTATATACCAAGGGGATACAA---CCGTTTGGC 602
Db 1067 SerProLeuGlyPheIleCys-----HisCysProLysGlyPheGluGlyProThrCys 1084
QY 603 TCC 605
Db 1085 Ser 1085

RESULT 9
GAT6_CHICK
ID GAT6_CHICK STANDARD; PRT; 387 AA.
AC PA3693;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription factor GATA-6 (GATA binding factor-6).
GN GATA6.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxId=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=White Leghorn;
RX MEDLINE=94365018; PubMed=8083222;
RA Laverriere A.C., Macneill C., Mueller C., Poelmann R.E.,
RA Burch J.B.E., Evans T.;
RT "GATA-4/6, a subfamily of three transcription factors transcribed
in developing heart and gut."
RL J. Biol. Chem. 269:23177-23184 (1994).
CC - SUBCELLULAR LOCATION: Nuclear.
CC - TISSUE SPECIFICITY: MORE ABUNDANT IN STOMACH, AND IN SMALL
INTESTINE. LOWER LEVELS IN LUNG, LIVER, OVARY AND HEART.
CC - SIMILARITY: CONTAINS 2 GATA-TYPE ZINC FINGERS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; U11889; AAA57505.1; -.
DR HSSP; P17679; 1GNF.
DR TRANSFAC; T02684; -.
DR InterPro; IPR000679; Znf_GATA.
DR InterPro; IPR001164; hrip_like.
DR Pfam; PF00320; GATA_2.
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; Znf_GATA_2.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 2.
DR PROSITE; PS50114; GATA_ZN_FINGER_2; 2.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
KM Nuclear protein.
FT ZN_FING 181 205 GATA-TYPE 1.
FT ZN_FING 235 259 GATA-TYPE 2.
FT DOMAIN 240 244 POLY-THR.
SQ SEQUENCE 387 AA; 40249 MW; 6EEF457BB34DB2C0 CRC64;

```

```

DB: 1 Gaps: 11
US-09-809-545a-1_copy_535_1143 (1-609) x GAT6_CHICK (1-387)
QY 27 CTACACCAATGGCGGAAATTAATCCAGTTGGGCGCGGTACAGCCCGCACTTCTA 86
Db 1 MethylGlnThrLeuAlaIleSerAlaSerGlnGly----- 12
QY 87 TGCAGGACGCGCTGTTGTGCGCAAGCCAAACAGAGGATCTTCATGACAGTGGCC 146
Db 13 -----ProIaProtyrApeGlySerProGlyGly---Phe-MethHisSerAlaPr 28
QY 147 CAGTTCCTGTATATCT-----TTCGAATGCTGCTTCCATA 188
Db 28 oSerSerProValItyrValProThrThrArgValGlySerValIleuProThrIleuProTy 48
QY 189 T----- 189
Db 48 rLeuGlnGlyGlyAlaAlaGlnProGlyHisAlaProIaGlyHisValItyrSerGI 68
QY 190 -CCGCGCGCCACTGCTGCAGCT-----GCATACCGAGGGCTTCACCTGAGCGCGTGG 242
Db 68 nProIaIaIaGlySerProSerItyrGlyAlaIaIaGlyAlaHisProSerGlyArg-- 87
QY 243 TCGCACCGGTACACACTTCAGAGCTCGCGCGGCCCAACCCCAATC----- 291
Db 88 -----PheProtyrSerAlaIaSerProProValAlaIaSerGlyAlaIa 101
QY 292 -----CCGCGCCATGCGCGGAGTATGTCATCAAGCCAGCTGATATGCG-----AATAAT 341
Db 101 rArgGlnGlnItyrGlyGlyGlyLeuAlaIaArgGlnGlnItyrAlaIeProIaPr 121
QY 342 GCTACAGGCGTGTACCTGCATACCGCTACCGCCAGCCACCCCTCCACTGCTGCTGC 401
Db 121 oLeuAlaGlySerItyr-----ProIaProtyrAlaIaIa 132
QY 402 CTAC--AGTGACATTACGACGACGAGTTATGCTCCGAGCCCTACACACACACTT-- 456
Db 132 rTyValGlyProGlnIleuGlyProIaItyrProIaIaIaProIaGlnIleuSerVal 152
QY 457 -----GCTCCAGCCGCC 468
Db 152 uHisCysLeuGlnGlyAlaArgAlaIaIaProIaIlePro 163

RESULT 10
FXEL_HUMAN
ID FXEL_HUMAN STANDARD; PRT; 376 AA.
AC 000356; O75765;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Forkhead box protein F1 (Forkhead-related protein FKHL15) (Thyroid
transcription factor 2) (TF2-2).
GN FOXE1 OR FKHL15 OR TF2 OR TF2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Keratinocytes;
RX MEDLINE=97312696; PubMed=9169137;
RA Chadwick B.P., Obermayr F., Frieschuf A.-M.;
RT "FKHL15, a new human member of the forkhead gene family located on
chromosome 9q22."
RL Genomics 41:390-396 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Macchia P.E., Mattei M.-G., Lapi P., Penzi G., Di Lauro R.;
RT "Cloning, chromosomal localization and identification of two
RT polymorphisms in the human thyroid transcription factor 2 gene
RT (TF2)."
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.

```

Alignment Scores:

Pred. No.:	1.07	Length:	387
Score:	99.00	Matches:	53
Percent Similarity:	35.94%	Conservative:	16
Best Local Similarity:	27.60%	Mismatches:	50
Query Match:	8.51%	Indels:	73

```

RN [3]
RP VARIANT BAMFORTH-LAZARUS SYNDROME VAL-65.
RX MEDLINE=98361170; PubMed=9697705;
RA Clifton-Bligh R.J., Wentworth J.M., Heinz P., Crisp M.S., John R.,
RA Lazarus J.H., Ludgate M., Chattejee V.K.;
RT "Mutation of the gene encoding human TTF-2 associated with thyroid
RT agenesia, cleft palate and choanal atresia.";
RL Nat. Genet. 19:399-401(1998).
CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR. COULD BE INVOLVED IN
CC THYROID GLAND ORGANOGENESIS.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- TISSUE SPECIFICITY: DETECTED IN ADULT BRAIN, PLACENTA, LUNG,
CC LIVER, SKELETAL MUSCLE, KIDNEY, PANCREAS, HEART, COLON, SMALL
CC INTESTINE TESTIS AND THYMUS. EXPRESSION WAS STRONGEST IN HEART AND
CC PANCREAS.
CC -1- DISEASE: DEFECTS IN FOXE1 ARE THE CAUSE OF BAMFORTH-LAZARUS
CC SYNDROME. A DISEASE ASSOCIATED WITH THYROID AGENESIS, CLEFT PALATE
CC AND CHOANAL ATRESIA.
CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U89995; AAC51294.1; -.
CC EMBL; Y1386; CAAT3816.1; -.
CC DR HSSP; 063245; 2HFM.
CC TRANSFAC; T02782; -.
CC DR Genew; HGNC:3806; FOXE1.
CC MIM; 602617; -.
CC DR InterPro; IPR001766; TF_Fork_head.
CC Pfam; PF00250; Fork_head; 1.
CC DR PRINTS; PR00053; FORKHEAD.
CC DR Prodom; PD000425; TF_Fork_head; 1.
CC DR SMART; SM00339; FH; 1.
CC DR PROSITE; PS00657; FORK_HEAD_1; 1.
CC DR PROSITE; PS00658; FORK_HEAD_2; 1.
CC DR PROSITE; PS00659; FORK_HEAD_3; 1.
CC KM DNA-binding; Nuclear protein; transcription regulation;
CC Disease mutation.
CC FT DOMAIN 7 12 POLY-PRO.
CC FT DNA_BIND 52 143 FORK-HEAD.
CC FT DOMAIN 164 182 POLY-ALA.
CC FT VARIANT 65 65 A -> V (IN BAMFORTH-LAZARUS SYNDROME).
CC FT CONFLICT 178 182 FTIDVAR_008857.
CC SQ SEQUENCE 376 AA; 38289 MW; 766534A7331572C CRC64;
Alignment Scores:
Score.No.: 1 18 Length: 376
Percent Similarity: 98.50 Matches: 45
Best Local Similarity: 36.48 Conservative: 13
Query Match: 28.30 Mismatches: 68
DB: 8.46 Indels: 33
Gaps: 6
US-09-809-545A-1_COPY_535_1143 (1-609) x FxE1_HUMAN (1-376)
OY 34 AATGCGTGAATTAATCATCAGTTGTGGCGCGGTACAGCCCGCACTTCTATGACGC 93
DB 125 AAGTCTTTPAlaleuAspPröAenAlaGlUAspMet-----PheGluSerGly 140
OY 94 ACGGCTCTGTGTGCCAGGCAACAGAGAGATCTTCATGATGACGTGCGCCCACTTCA 153
DB 141 SerPheLeuArgArgGlybArGpPheGlybArSerAspLeu----- 154
OY 154 CTTGTATATACCTTCGCAATCCTGCTGCTTCATATCCGCGCCCACTGTGACGTGCA 213

```

```

Db 155 -----SerThyTyrProAlaTyrMetHisAspAlaAlaAlaAlaAla 170
Oy 214 TACCGAGGGGCTACCTTCGAGGCGCGTGTGGACACCGGTGTCAACACCTTCAGAGCTG 273
Db 171 AlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 190
Oy 274 GCGCCGCCACCCCGCCATCCCGGCTATGCGAGAGTGTATGTATCAAGAGCCAGTGATGAC 333
Db 191 AlaArgProPro-----TyrProGlyAlaValTyr----- 200
Oy 334 AATAAATGCTACAGGGGTGTTCAGCTTCATACCGCTACGCCACGCCACCCCTGCACT 393
Db 201 -----AlaGlyTyrAlaProProSerLeuAlaAlaProProProValTyr 215
Oy 394 GCTGCTGCTTACAGTACGACAGTTCACGAGCAGACTTATGCT-----GCCGACCCCTAC 444
Db 216 TyrProAlaAlaSerProGlyProCysArgValAPheGlyLeuValProGlyArgProLeu 235
Oy 445 CACCAACACACTGTGCTCCAGCCCGCCACCTACGAGCGGTGTGCGCATGATGCTTTTGGC 501
Db 236 SerProGlyLeuGlyProAlaProSer---GlyProGlyGlySerCysAlaPheAla 253

RESULT 11
EWS_MOUSE
ID EWS_MOUSE STANDARD; PRT; 655 AA.
AC 061545;
AD 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE RNA-binding protein EWS.
GN EWS OR EWSH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RS SEQUENCE FROM N.A.
RC Tissue=Testis;
RX MEDLINE=95130099; PubMed=7829990.
RA Plougaard B., Mattei M.-G., Thomas G., Delattre O.;
RT "Cloning and chromosome localization of the mouse Ews gene.";
RL Genomics 23:278-281(1994).
CC -1- FUNCTION: MIGHT FUNCTION AS A REPRESSOR (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNM).
CC -1- SIMILARITY: CONTAINS 1 RANBP2-TYPE ZINC FINGER.
CC -1- SIMILARITY: BELONGS TO THE TEST FAMILY OF RNP PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X79233; CAA55815.1; -.
DR MGD; MGI:99960; Ewhh.
DR InterPro; IPR000504; RNA_rec_mot.
DR InterPro; IPR001876; Znf_RanBP.
DR Pfam; PF00076; Rm; 1.
DR Pfam; PF00641; Zf-RanBP; 1.
DR SMART; SM00360; RRM; 1.
DR SMART; SM00547; ZNF_RBZ; 1.
DR PROSITE; PS50106; IQ; FALSE_NEG.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
DR PROSITE; PS01358; ZF_RANBP2_1; 1.
DR PROSITE; PS50199; ZF_RANBP2_2; 1.
KW Nuclear protein; zinc-finger; zinc; RNA-binding; Metal-binding;
KW Calmodulin-binding; Repeat; Methylation; Phosphorylation;
KW Transcription regulation; Repressor.

```

FT	DOMAIN	1	285	EAD (GLN/PRO/THR-RICH) .
FT	DOMAIN	256	285	IO.
FT	DOMAIN	360	446	RNA-BINDING (RRM) .
FT	ZN_FING	517	548	RANBP2-TYPE.
FT	DOMAIN	8	285	31 X APPROXIMATE TANDEM REPEATS.
FT	REPEAT	8	16	1.
FT	REPEAT	17	27	2.
FT	REPEAT	28	34	3.
FT	REPEAT	35	42	4.
FT	REPEAT	43	50	5.
FT	REPEAT	51	59	6.
FT	REPEAT	60	68	7.
FT	REPEAT	69	75	8.
FT	REPEAT	76	84	9.
FT	REPEAT	85	91	10.
FT	REPEAT	92	110	11.
FT	REPEAT	111	116	12.
FT	REPEAT	117	125	13.
FT	REPEAT	126	156	14.
FT	REPEAT	157	163	15.
FT	REPEAT	164	170	16.
FT	REPEAT	171	177	17.
FT	REPEAT	178	188	18.
FT	REPEAT	189	193	19.
FT	REPEAT	194	201	20.
FT	REPEAT	202	206	21.
FT	REPEAT	207	212	22.
FT	REPEAT	213	218	23.
FT	REPEAT	219	224	24.
FT	REPEAT	225	230	25.
FT	REPEAT	231	238	26.
FT	REPEAT	239	245	27.
FT	REPEAT	246	252	28.
FT	REPEAT	253	259	29.
FT	REPEAT	260	276	30.
FT	REPEAT	277	285	31.
FT	DOMAIN	300	339	ARG/GLY/PRO-RICH.
FT	DOMAIN	303	312	ARG/GLY/PRO-RICH.
FT	DOMAIN	358	639	ARG/GLY/PRO-RICH.
FT	MOD_RES	300	300	(BY SIMILARITY) .
FT	MOD_RES	302	302	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	304	304	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	309	309	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	314	314	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	317	317	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	321	321	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	454	454	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	463	463	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	470	470	METHYLATION (MONO- OR DI-) (BY SIMILARITY) .
FT	MOD_RES	489	489	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	493	493	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	499	499	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	502	502	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	505	505	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	562	562	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	564	564	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	571	571	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	574	574	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	580	580	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	588	588	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	591	591	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	595	595	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	599	599	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	602	602	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	606	606	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	614	614	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	632	632	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	635	635	METHYLATION (DI-) (BY SIMILARITY) .
FT	MOD_RES	266	266	PHOSPHORYLATION (BY PKC) (BY SIMILARITY)
QO	SEQUENCE	655 AA;	68418 MW;	50735EDB54247D69 CRC64;

Alignment Scores:

Pred. No.:	1.2	Length:	655
Score:	98.50	Matches:	49
Percent Similarity:	34.88%	Conservative:	11
Best Local Similarity:	28.49%	Mismatches:	71
Query Match:	8.46%	Indels:	41
DB:	1	Gaps:	9

US-09-809-545A-1_COPY_535_1143 (1-609) x EMS_MOUSE (1-655)	
QY 85	TATGAGCAGCAGCGTGTGTGTGTCAGGCC--AACGAGGAGATCTTCATGTCAGCT 141
DB 29	TyrAlaGlnThr-----GlnAlaTyrGlnGlnSerTyrGlnTyrGly 45
QY 142	GGCCCCAGTTACCTGTATATACCTTTCGCAATGCTGCTGCTTCATGTCGGCCGCACT 201
DB 46	GlnProThrAspValSerTyrThrGlnAlaGlnThrThrAlaThrTyrGlnThrAla 65
QY 202	GCTGAGAGTCGATACCGAGGGGCTCACCTTCGAGCCGCTGTCGACCCGTCGACACAC 261
DB 66	TyrAlaThrSerTyr-----GlyGlnProProThrGlnTyrSerThr 79
QY 262	TTGAGAGCTGGGGGCCCCCAGCCCAATCCGGCTATGGCGGAGTAGTCATCAAG 321
DB 80	-----ProThrAlaProGln-----AlaTyrSerGln 88
QY 322	CCAGGTATGTCATTAATTGCTACAGGGTGTACCGTCGATACCGCTACGCCACCC 381
DB 89	ProValGlnGlyTyr-----GlyThrGlyAlaTyrAspSerThrAla 103
QY 382	ACCCTGTCAGTCGTGCTGCTTACAGTGACAGTACGAGACGATTATGCTGCCAGCCC 441
DB 104	ThrValThrThrThrGlnAla-----SerTyrAlaAlaGlnSer 116
QY 442	TACCAACACACACTGCTGCTCCAGCCCCACCTACGCGCTGTGTCATGAATGCTTTGCG 501
DB 117	AlaTyrGlnTyrGlnProAlaTyrProThrTyrGln-----Gln 130
QY 502	CCCTTGACCATGCCAGACTAGAGCCATGCTATATGCTGCTCCGTTCTTTCTCA 561
DB 131	ProThrAlaThrAlaProThrThrProGlnAspGlyAsnLysProAlaGlnThrSerGln 150
QY 562	TTGCAGGCTAGTATATACCAAGGGGATACAAACCT 597
DB 151	ProGlnSerSer-----ThrGlyGlnTyrAsnGln 160

RESULT 12	
HMEV_DROME	STANDARD; PRT; 376 AA.
AC P06602; P07667; Q9VSE6;	
DT 01-JAN-1988 (Rel. 06, Created)	
DT 16-OCT-2001 (Rel. 40, Last sequence update)	
DT 15-JUN-2002 (Rel. 41, Last annotation update)	
DE Segmentation protein even-skipped.	
OS EVE OR CG3328.	
OS Drosophila melanogaster (Fruit fly).	
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;	
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;	
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
OX NCBI_TaxId=7227;	
XP [1]	
RP SEQUENCE FROM N.A.	
RA MEDLINE=87051744; PubMed=2877745;	
RA McDonald P.M., Ingham P., Struhl G.;	
RT "Isolation, structure, and expression of even-skipped: a second pair	
RT rule gene of Drosophila containing a homeo box.";	
RL Cell 47:721-734(1986).	
RN [2]	
RP SEQUENCE FROM N.A.	
RA MEDLINE=87218536; PubMed=2884106;	
RA Frasch M., Hoey T., Rushlow C., Doyle H., Levine M.;	
RT "Characterization and localization of the even-skipped protein of	
RT Drosophila.";	
RL EMBO J. 6:749-759(1987).	

[3]
RP SEQUENCE FROM N.A..
RC STRAIN=Berkeley;
RX MEDLINE=20196406; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bandaru D., Bolshakov S.,
RA Butkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Butts K.C., Busen D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferriz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heitman T.J., Hernandez J.R., Houck J.,
RA Hootin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam G.,
RA Jacot J.M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodali C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,
RA Palazzolo M., Peltom G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Stimpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svaytshes R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodge T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RT Science 287:2185-2195 (2000).
RN [4]
RP SEQUENCE OF 1-58 FROM N.A..
RC STRAIN=OR-RC, WA-F, ZMS6, AF-S, and FL-S;
RX MEDLINE=96038621; PubMed=8524036;
RA Ludwig M.Z., Kretzman M.;
RT "Evolutionary dynamics of the enhancer region of even-skipped in
RT *Drosophila*.";
RL Mol. Biol. Evol. 12:1002-1011 (1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 70-189.
RX MEDLINE=96134926; PubMed=8557047;
RA Hirsch J.A., Aggarwal A.K.;
RT "Structure of the even-skipped homeodomain complexed to AT-rich DNA:
RT new perspectives on homeodomain specificity.";
RL EMBO J. 14:6280-6291 (1995).
CC -1- FUNCTION: MAY PLAY A ROLE IN DETERMINING NEURONAL IDENTITY. MAY BE
CC DIRECTLY INVOLVED IN SPECIFYING IDENTITY OF INDIVIDUAL NEURONS.
CC PAIR-RULE PROTEIN REQUIRED FOR SEGMENTATION; INVOLVED IN
CC TRANSFORMING THE BROAD, SPATIAL, APERIODIC EXPRESSION PATTERNS OF
CC THE GAP GENES INTO A SYSTEM OF PRECISE PERIODIC EXPRESSION
CC PATTERNS OF THE PAIR-RULE AND SEGMENTARY POLARITY GENES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE EVEN-SKIPPED HOMEOBOX FAMILY.
CC PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

```

CC      or send an email to license@sb-sib.ch).
-----
DR      EMBL, M14767; AAA28522.1; -.
DR      EMBL, X05138; CA28784.1; -.
DR      EMBL, X05138; CA28785.1; ALT SEQ.
DR      EMBL, AE001831; AAF58865.1; -.
DR      EMBL, U32087; AAB05358.1; -.
DR      EMBL, U32088; AAB05359.1; -.
DR      EMBL, U32089; AAB05360.1; -.
DR      EMBL, U32090; AAB05361.1; -.
DR      EMBL, U32091; AAB05362.1; -.
DR      PIR, A26066; A26066.
DR      HSSP, P14653; 1B72.
DR      TRANSFAC, T00272; -.
DR      FLYbase; FBgn000606; eve.
DR      InterPro; IPR000047; HTH repressor.
DR      InterPro; IPR001356; Homeobox.
DR      Pfam; PF00046; homeobox; 1.
DR      PRINTS; PR00024; HOMEBOX.
DR      PRINTS; PR00031; HTHREPRESSR.
DR      ProDom; PD000010; Homeobox; 1.
DR      SMART; SMO0389; HOX; 1.
DR      PROSITE; PS00027; HOMEBOX_1; 1.
DR      PROSITE; PS50071; HOMEBOX_2; 1.
DR      DNA-binding; Developmental protein; Embryo; Pair-rule protein;
KW      Transcription regulation; Homeobox; Nuclear protein.
FT      DNA BIND       70      129      HOMEBOX.
FT      DOMAIN         165      179      ALA-RICH.
FT      CONFLICT       300      300      L -> V (IN REF. 1).
SQ      SEQUENCE       376 AA; 39970 MW; 59058E0634B1BED0 CRC64;

```

Alignment Scores:		1.3	Length:	376
Pred. No.:		98.00	Matches:	45
Score:		28.65%	Conservative:	8
Percent Similarity:		24.32%	Mismatches:	36
Best Local Similarity:		8.42%	Indels:	76
Query Match:			Gaps:	7
DB:		1		

US-09-809-545A-1_COPY_535_1143 (1-609) x HMEV_DROME (1-376)	
Qy	64 GCGGCTACAGCCCCCGACTTCTATGACAGGACGGTGTCTTTGTGCCAGGCCACACGAG 123
Dp	138 AlaValTyrSerAspProAlaPheAlaAlaSerIleLeuGlnAlaAlaAlaAsnSerVal 157
Qy	124 GGATTTCCAGTATACAGTGGCCCGACTTACTTGTATATCTTGCATGCCCGGCTTT 183
Dp	158 Gly-----MetProTyrPro 162
Qy	184 CCATAT---CCGGCGGCACACTCTGACAGCTGCATACCGAGGGGCTCACTTCGAGGGCGT 246
Dp	163 ProTyrAlaProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 177
Qy	241 GGTGGCACCGGTACACACCTTCAGAGCTGGCGGCCGCCACCC----- 285
Dp	178 -----ValAlaThrAsnProMetMetCysIleThrGlyMetProProMetGlyMetProGln 195
Qy	286 -----CCAAATCCGGGCGCTATGGCGGAGTGAAGTATCAAGAGCCAGTATATGGC 333
Dp	196 MetProThrMetGlnMetProGlyHisSerGlyHisAlaGlyHisProSerProTyrGly 215
Qy	334 AATAAATTGCTACAGGGGTGTAGCGTGCATACCGCTACCGCCACGCCACCCCTGCAGT 393
Dp	216 Gln-----TyrAlaGlyThr----- 220
Qy	394 GCTGCTGCCCTACAGTACGACGTTACGAGCAGAGTTATGCTGCCGACCCCTACACACACA 455
Dp	221 -----ProTyrHisIlePro 225
Qy	454 CTTGCTCAGACCCCACTACAGGGGTGTGGTCATGAATGCTTTTGGGCCCTTGACGAT 511
Dp	226 AlaArgProAlaProProHisIleGly----- 233

Qy	514	GC	CAAGACTAGAGCATCGATCGATGGCTTCGTCCTTCATTCGACGCTAGT	573
Dy	236	-----	ProhibitinshsProhibitinMeuMetClySerSerAlaThrGlySer	251
Dy	574	ATATACCAAGGGCGA	588	
Dy	252	SerTyrSerAlaGly	256	
RESULT 13				
BRN1_RAT	ID	BRN1_RAT	STANDARD:	PRT; 497 AA.
AC	Q63262;			
DT	15-JUL-1998	(Rel. 36, Created)		
DT	15-JUN-2002	(Rel. 38, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	Brain-specific Homeobox/POU domain protein 1 (BRN-1 protein).			
GN	POU3F3 OR BRN1 OR BRN-1 OR RH52.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxId=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98070400; PubMed=9405434;			
RA	Schreiber J., Enderich J., Sock E., Schmidt C., Richter-Jansberg C.,			
RA	Wegner M.			
RT	"Redundancy of class III POU proteins in the oligodendrocyte			
RT	lineage.";			
RL	J. Biol. Chem. 272:32286-32293(1997).			
RN	[2]			
RP	SEQUENCE OF 325-449 FROM N.A.			
RC	TISSUE=Hypothalamus;			
RX	MEDLINE=92228769; PubMed=1348858;			
RA	le Moine C., Young W.S.;			
RT	"RH52, a POU domain-containing gene, and its expression in developing			
RT	and adult rat.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 89:3285-3289(1992).			
CC	-1- SUBCELLULAR LOCATION: Nuclear (By similarity).			
CC	-1- TISSUE SPECIFICITY: BRAIN.			
CC	-1- DEVELOPMENTAL STAGE: EXPRESSED FROM EMBRYONIC DAY 11.5 INTO			
CC	ADULTHOOD.			
CC	-1- SIMILARITY: BELONGS TO THE POU TRANSCRIPTION FACTOR FAMILY.			
CC	-1- SIMILARITY: CONSTRAINTS 1 HOMEBOX DOMAIN.			
CC	CLASS-3 SUPERFAMILY.			
CC	-1- This SWISS-PROT entry is copyright. It is produced through a collaboration -			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use. By non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AJ001641; CAA04893.1; -.			
DR	EMBL; M84644; AAA42041.1; -.			
DR	HSSP; P14859; 1OCT.			
DR	InterPro; IPR001356; Homeobox.			
DR	InterPro; IPR000327; POU domain.			
DR	Pfam; PF00046; homeobox_1.			
DR	Pfam; PF00157; pou; 1.			
DR	PRINTS; PR00028; POUDOMAIN.			
DR	ProDom; PD000010; Homeobox; 1.			
DR	ProDom; PD000583; POU domain; 1.			
DR	SMART; SM00389; HOX; 1.			
DR	SMART; SM00352; POU; 1.			
DR	PROSITE; PS00027; HOMEBOX_1; 1.			
DR	PROSITE; PS00071; HOMEBOX_2; 1.			
DR	PROSITE; PS00035; POU_1; 1.			
DR	PROSITE; PS00465; POU_2; 1.			
KW	Nuclear protein; DNA-binding; Homeobox.			
FT	DOMAIN 28 49			
FT	POLY-GLY.			
FT	DOMAIN 103 114			
FT	POLY-ALA.			
FT	DOMAIN 135 143			
FT	POLY-PRO.			

FT	DOMAIN	173	178	POLY-PRO.
FT <td>DOMAIN</td> <td>188</td> <td>203</td> <td>POLY-ALA.</td>	DOMAIN	188	203	POLY-ALA.
FT	DOMAIN	238	249	POLY-GLY.
FT	DOMAIN	269	280	POLY-HIS.
FT	DOMAIN	294	301	POLY-GLY.
FT	DOMAIN	315	385	POU.
FT	DNA BIND	403	462	HOMEBOX.
SEQ	SEQUENCE	497 AA;	50226 MM;	0064050E343ABC2 CRC64;

Alignment Scores:			
Pred. No.:	1.58	Length:	497
Score:	97.00	Matches:	53
Percent Similarity:	32.17%	Conservative:	21
Best Local Similarity:	23.04%	Mismatches:	76
Query Match:	8.33%	Indels:	80
DB:	1	Gaps:	11

US-09-809-545A-1_COPY_535_1143 (1-609) x BRN1_RAT (1-497)
QY 55 GTTGGGGGGGGGCTACAGCCCGCATCTTATGACGACGAGCGCTGTGTGCCAGGCC 1144
DB 148 VallysglglValIglYarglualsplehlsIalglYthralalau 1633
QY 115 AACGAGAGGAGTCTTCATGATACAGGCGGCCGTTCACTGTATATACTTGTGCAATG 1744
DB 164 -----HshlsarGlYProProhIsleuGlYProProProPro 1777
QY 175 CCTGGCTTTCCATATCCG-----GCCGCGACTGTGCA 207
DB 178 ProhIsnglnghIaProglYglYtrpGlYalalalalalalalalalalalalalalal 1977
QY 208 GCTGCATATCCGAGGGGCTCACTTCGAGCGCGTGTGCGACCGTGTACAAACCTTGAGA 267
DB 198 Alalalalalalalalalalalshlsleu-----ProserMeta 209
QY 268 GCTGGGGGGGGCCCGCCCGCATCCGGCGCATGGCGGAGAGTGTACAAAGCCAGTGC 327
DB 210 glYglYnglnghIaProProPro-----GlnserleuGlYserGlnPro 224
QY 328 TATGGCAATTAATTGTCACAGGT----- 351
DB 225 ---glYglYrPhethrValaenGlYmetleuserlaIaProProglYProglYglYglY 243
QY 352 -----GGTACCGTGCATACCGGTACGCCCGCCAGCC 384
DB 244 glYglYalaglYglYglYalaglInserleuValhIsProglYleuValaXglYasPrh 263
QY 385 CCTGCACCTGCTGCTGCTACAGTACAGATTAACGAGATTAAGTGTGCCGACCCCTAC 444
DB 264 ProgluleuIa 283
QY 445 ---CACACACACTTGGCTCCAGCCCGCCACCGACCGCTGTGT-----GCCATG 489
DB 284 ProhIshtlaIa 302
QY 490 AATGCTTTTGGCGCCTTGACCGATGCCAAGACTAGGAGCCATGCTGAT----- 537
DB 303 AsnserhtIsaerProhIsaerPaerGlualsPrhrothserPaerleuGlInPhe 322
QY 538 -----GATGGGGGTCTC 549
DB 323 AlaYsglnPheYsglnaYarglIeYlsleuGlYrPhethrGlnalaaIaIaIaIaIaIaIa 342
QY 550 GTTCTTTTTCATTCAGCAGGCTAGATATAC 579
DB 343 AlaIeuglYlnrleuYrGlYashtValaPhe 352

RESULT 14			
ID	PC15_MOUSE	STANDARD;	PRT; 1943 AA.
AC	Q99P01;		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		

PT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Protocadherin 15 precursor.	
GN	PCDH15.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_Taxid=10090;	
PN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEBLIN=20578757; PubMed=1138007;	
RA	Alegreman K.N., Murcia C.L., Kwon H.Y., Pawlowski K.S., Wright C.G.,	
RT	Moylechik R.P.;	
RL	"The mouse Ames waltzer hearing-loss mutant is caused by mutation of	
RT	pcdh15, a novel protocadherin gene."	
RL	Nat. Genet. 27:99-102(2001).	
RN	[2]	
RP	TISSUE SPECIFICITY.	
RX	MEBLIN=21322691; PubMed=11429292;	
RA	Murcia C.L., Moylechik R.P.;	
RT	"Expression of Pcdh15 in the inner ear, nervous system and various	
RL	epithelia of the developing embryo."	
CC	Mech. Dev. 105:163-166(2001).	
CC	-1- FUNCTION: Calcium-dependent cell-adhesion protein. Essential for	
CC	maintenance of normal retinal and cochlear function. Required for	
CC	inner ear neuroepithelial cell elaboration.	
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).	
CC	-1- TISSUE SPECIFICITY: Expressed in brain and sensory epithelium of	
CC	the developing inner ear. Also found in the spleen, developing	
CC	eye, dorsal root ganglion, dorsal aspect of neural tube, floor	
CC	plate and endoplasmic reticulum cells adjacent to the neural canal.	
CC	-1- DEVELOPMENTAL STAGE: Highest level of expression is detected at	
CC	embryonic day 16.	
CC	-1- DISEASE: Defects in PCDH15 are the cause of the Ames waltzer (av)	
CC	phenotype. It is characterized by deafness and a balance disorder,	
CC	associated with the degeneration of inner ear neuroepithelia.	
CC	-1- SIMILARITY: CONTRAINS 11 CADHERIN DOMAINS.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
CC	EMBL; AF281899; AAG53891.1; -	
DR	HSSP; P09803; ISUH.	
DR	MGP; MGI:1891428; Pcdh15.	
DR	InterPro; IPR002126; Cadherin.	
DR	PIfam; PF00028; cadherin; 10.	
DR	PRINTS; PR00205; CADHERIN.	
DR	SMART; SM00112; CA; 11.	
DR	PROSITE; PS00332; CADHERIN 1; 4.	
DR	PROSITE; PS50268; CADHERIN 2; 11.	
DR	Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;	
KW	Signal; Multigene family; Vlsion.	
FT	SIGNAL	1
FT	CHAIN	27
FT	DOMAIN	27 1381
FT	TRANSMEM	1382 1402
FT	DOMAIN	1403 1993
FT	DOMAIN	45 152
FT	DOMAIN	153 270
FT	DOMAIN	283 400
FT	DOMAIN	401 514
FT	DOMAIN	515 621
FT	DOMAIN	622 722
FT	DOMAIN	724 824
FT	DOMAIN	825 931
FT	DOMAIN	932 1040
FT	DOMAIN	1042 1149
FT	DOMAIN	1150 1264
FT	DOMAIN	1437 1448
FT	POLY-PRO.	

FT	DOMAIN	1772	1778	POLY.-PRO.		(POTENTIAL).
FT	CARBOHYD	1804	1812	N-LINKED (GLCNAC .)	.	(POTENTIAL).
FT	CARBOHYD	57	57	N-LINKED (GLCNAC .)	.	(POTENTIAL).
FT	CARBOHYD	102	102	N-LINKED (GLCNAC .)	.	(POTENTIAL).
FT	CARBOHYD	206	206	N-LINKED (GLCNAC .)	.	(POTENTIAL).
FT	CARBOHYD	424	424	N-LINKED (GLCNAC .)	.	(POTENTIAL).
FT	CARBOHYD	564	564	N-LINKED (GLCNAC .)	.	(POTENTIAL).
FT	CARBOHYD	667	667	N-LINKED (GLCNAC .)	.	(POTENTIAL).
FT	CARBOHYD	729	729	N-LINKED (GLCNAC .)	.	(POTENTIAL).
FT	CARBOHYD	773	773	N-LINKED (GLCNAC .)	.	(POTENTIAL).
FT	CARBOHYD	826	826	N-LINKED (GLCNAC .)	.	(POTENTIAL).
FT	CARBOHYD	856	856	N-LINKED (GLCNAC .)	.	(POTENTIAL).
FT	CARBOHYD	1069	1069	N-LINKED (GLCNAC .)	.	(POTENTIAL).
FT	CARBOHYD	1089	1089	N-LINKED (GLCNAC .)	.	(POTENTIAL).
FT	CARBOHYD	1180	1180	N-LINKED (GLCNAC .)	.	(POTENTIAL).
SQ	SEQUENCE	1943 AA;	214816 MM;	E3JD7DB9F5F73B652 CRC64;		
 Alignment Scores:						
Pred. No.:		1.66	Length:	1943		
Score:		97.00	Matches:	16		
Percent Similarity:		38.69%	Conservative:	49		
Best Local Similarity:		27.38%	Mismatches:	65		
Query Match:		8.33%	Indels:	38		
DB:		1	Gaps:	9		
 US-09-809-545A-1_COPY_535_1143 (1-609) x PC15_MOUSE (1-1943)						
Oy	59 TGGGCGGGG-----TCTACAGCCCGCACTTCATGCAGCAGCGGTGCTTTGCC	109				
Db	1672 TyperhisPheserPheserThrleuProthrIleserArglaValgluleuGlYser	1691				
Oy	110 AGGCCAACACAGAGGACTTCCA-----TGTAAGTAGGCCCGCCAGTTCATTGATATA	163				
Db	1692 GluProasnValValThrSerProllaalaaPyethrLeu-----	1704				
Oy	164 CTTCTGCATATGCTCGCTTCATATCCGCGCCACTCTGCAGCTCATACAGGAGGG	223				
Db	1705 -----GlueuserProProleuAlargProAlglleuAlnsr	1717				
Oy	224 CTCACCTTCGAGGCGGTGCGCACCGGTATACACACTTCAGAGCTCGGCGC-----	277				
Db	1718 LeuserSerlysaArgluThrProthrCysalaSerAspThglubProlysrghsnser	1737				
Oy	278 -----CCCCACCCCMAATC-----CGGCTATAGGCGGAGTAGTATCAAG	319				
Db	1738 PheglullealaproHisProProSerillephealProleuProhisarProleu	1757				
Oy	320 AGCCAGTGTATGCGAATAAATGTCTACAGGTGTTACGCTGCATACCGCTCGCCAGC	379				
Db	1758 ArgProprollealaePheThrTrpne-----ProleuProleuserProProasn	1774				
Oy	380 CCACCCCTGCACCTGCTGCTGCCTACAGTGACGATTCAGGAGAGTTATGCTGCCGACC	439				
Db	1775 ProProProProgluInleu-----ValThrPheSerleuProIleserThr	1789				
Oy	440 CCTACACACACACACTTGTCTCCAGCCGCCACCTACAGCGGTGTGCCATGAATGCTTTG	499				
Db	1790 PropProThrSerSerleuProleuProProProleuserleuProPro-----ProPro	1807				
Oy	500 CGCCCTTGACCGATGCCAGAAGTAA 523					
Db	1808 ArgPro--ProAlaProArgleu 1814					
 RESULT 15						
ID	NTC4_MOUSE	STANDARD:	PRT:	1964 AA.		
AC	P31695; Q62389; Q62390; Q35442; Q9N1W; O88314; O88316; Q9R1X0;					
DT	01-JUN-1993 (Rel. 26, Created)					
DT	01-NOV-1997 (Rel. 35, Last sequence update)					
DT	15-JUN-2002 (Rel. 41, Last annotation update)					
DE	Neurogenic locus notch homolog protein 4 precursor (Notch 4)					
GN	(Contains: Transforming protein Int-3. NOTCH4 OR INT3 OR INT-3.					

OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP MEDLINE=92194507; PubMed=1312643;
RA Robbins J., Blondel B.J., Callahan D., Callahan R.;
RT "Mouse mammary tumor gene int-3: a member of the notch gene family
RT transforms mammary epithelial cells.";
RL J. Virol. 66:2594-2599(1992).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RA MEDLINE=97294599; PubMed=9150355;
RX Gallahan D., Callahan R.;
RT "The mouse mammary tumor associated gene INT3 is a unique member of
RT the NOTCH gene family (NOTCH4).";
RL Oncogene 14:1883-1890(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung, and Testis;
RX MEDLINE=96281668; PubMed=8681805;
RA Uytendaele H., Marazzi G., Wu G., Yan Q., Sassoon D., Kitajewski J.;
RT "Notch4/int-3, a mammary proto-oncogene, is an endothelial
RT cell-specific mammalian Notch gene.";
RL Development 122:2251-2259(1996).
RN [4]
RP SEQUENCE FROM N.A.
RA Bowen L., Mahairas G., Qin S., Ahearn M.E., Dankers C., Laaky S.,
RA Loretz C., Schmidt S., Tipton S., Traicoff R., Zackone K., Hood L.;
RT "Sequence of the mouse major histocompatibility locus class III
RT region.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1436-1600 FROM N.A.
RX MEDLINE=99252212; PubMed=10233982;
RA Lee J.-S., Haruna T., Ishimoto A., Honjo T., Yanagawa S.-I.;
RT "Intracisternal type A particle-mediated activation of the Notch4/int3
RT gene in a mouse mammary tumor: generation of truncated Notch4/int3
RT mRNAs by retroviral splicing events.";
RL J. Virol. 73:5166-5171(1999).
RN [6]
RP FUNCTION.
RX MEDLINE=21244657; PubMed=11344305;
RA Uytendaele H., Ho J., Rossant J., Kitajewski J.;
RT "Vascular patterning defects associated with expression of activated
RT Notch4 in embryonic endothelium.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:5643-5648(2001).
RN [7]
RP SEQUENCE OF 1463-1964, POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS
RP OF VAL-1463.
RX MEDLINE=21523956; PubMed=11518718;
RA Saxena M.T., Schroeder E.H., Mumm J.S., Kopan R.;
RT "Murine notch homologs (N1-4) undergo presenilin-dependent
RT proteolysis";
RL J. Biol. Chem. 276:40268-40273 (2001).
RN [8]
RP POST-TRANSLATIONAL PROCESSING.
RX MEDLINE=21374376; PubMed=11459941;
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT "Conservation of the biochemical mechanisms of signal transduction
RT among mammalian Notch family members.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
RN [9]
RP FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity). May regulate branching
CC morphogenesis in the developing vascular system.
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC terminal fragment N(IEC) which are probably linked by disulfide

CC bonds.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytic processing NICD is translocated to the nucleus.
CC -1- TISSUE SPECIFICITY: Highly expressed in lung, moderately in heart
CC kidney, and at lower levels in the ovary and skeletal muscle. A
CC very low expression is seen in the brain, intestine, liver and
CC testis.
CC -1- DEVELOPMENTAL STAGE: Highly expressed in endothelial cells during
CC embryonic development from 9.0 d.p.c.
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(IEC). Following
CC ligand binding, it is cleaved by TGF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane.
CC -1- PTM: Phosphorylated.
CC -1- DISEASE: Loss of the extracellular domain causes constitutive
CC activation of the Notch protein, which leads to hyperproliferation
CC of glandular epithelial tissues and development of mammary
CC carcinomas.
CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC -1- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M80456; AAB38377.1; -;
DR EMBL: U43691; AAC52630.1; -;
DR EMBL: U43691; AAC52631.1; -;
DR EMBL: AF030001; AAB82004.1; -;
DR EMBL: AB016771; BAA32281.1; ALT SEQ.
DR EMBL: AB016772; BAA32283.1; ALT INIT.
DR EMBL: AB016773; BAA32284.1; ALT INIT.
DR EMBL: AB016774; BAA32285.1; -;
DR PIR: A38072; TWMT3.
DR HSP: P08709; IBF9.
DR MGD: MGI:107471; Notch4.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000152; Asx hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR000800; Notch.
DR Pfam: PF00008; EGF; 27.
DR Pfam: PF00023; ank; 6.
DR Pfam: PF00066; notch; 2.
DR PRINTS: PRO1415; ANKYRIN.
DR PRINTS: PRO0010; EGFBLDOD.
DR PRINTS: PRO1452; NOTCH.
DR SMART: SM00248; ANK; 5.
DR SMART: SM00179; EGF_Ca; 11.
DR SMART: SM00001; EGF_like; 15.
DR SMART: SM00004; NL; 2.
DR PROSITE: PSS0088; ANK_REPEAT; 5.
DR PROSITE: PSS0297; ANK_REPEAT_REGION; 1.
DR PROSITE: PSS00010; ASX_HYDROXYL; 11.
DR PROSITE: PSS00022; EGF_1; 28.
DR PROSITE: PS01186; EGF_2; 21.
DR PROSITE: PS01187; EGF_CA; 9.
KW Receptor; Transcription regulation; Activator; Differentiation;

KM	Developmental protein; Repeat; ANK repeat; EGF-like domain;					
KW	Transmembrane; Glycoprotein; Signal; Phosphorylation; Proto-oncogene.					
FT	SIGNAL	1	20	POTENTIAL.		
FT	CHAIN	21	1964	NEUROSPIC LOCUS NOTCH HOMOLOG PROTEIN 4		
FT	CHAIN	1411	1964	TRANSFORMING PROTEIN INT-3.		
FT	CHAIN	1428	1964	NOTCH EXTRACELLULAR TRUNCATION.		
FT	CHAIN	1463	1964	NOTCH INTRACELLULAR DOMAIN.		
FT	DOMAIN	21	1443	EXTRACELLULAR (POTENTIAL).		
FT	TRANSMEM	1444	1464	POTENTIAL.		
FT	DOMAIN	1465	1964	CYTOPLASMIC (POTENTIAL)		
FT	DOMAIN	21	60	EGF-LIKE 1.		
FT	DOMAIN	61	112	EGF-LIKE 2.		
FT	DOMAIN	115	152	EGF-LIKE 3.		
FT	DOMAIN	153	189	EGF-LIKE 4.		
FT	DOMAIN	191	229	EGF-LIKE 5.		
FT	DOMAIN	231	271	EGF-LIKE 6.		
FT	DOMAIN	273	309	EGF-LIKE 7.		
FT	DOMAIN	311	350	EGF-LIKE 8.		
FT	DOMAIN	352	388	EGF-LIKE 9.		
FT	DOMAIN	389	427	EGF-LIKE 10.		
FT	DOMAIN	429	470	CALCIUM-BINDING (POTENTIAL)		
FT	DOMAIN	472	508	EGF-LIKE 11.		
FT	DOMAIN	510	546	EGF-LIKE 12.		
FT	DOMAIN	548	584	EGF-LIKE 13.		
FT	DOMAIN	586	622	EGF-LIKE 14.		
FT	DOMAIN	623	656	EGF-LIKE 15.		
FT	DOMAIN	658	686	EGF-LIKE 16.		
FT	DOMAIN	688	724	EGF-LIKE 17.		
FT	DOMAIN	726	762	EGF-LIKE 18.		
FT	DOMAIN	764	800	EGF-LIKE 19.		
FT	DOMAIN	803	839	EGF-LIKE 20.		
FT	DOMAIN	841	877	EGF-LIKE 21.		
FT	DOMAIN	878	924	EGF-LIKE 22.		
FT	DOMAIN	926	962	EGF-LIKE 23.		
FT	DOMAIN	964	1000	EGF-LIKE 24.		
FT	DOMAIN	1002	1040	EGF-LIKE 25.		
FT	DOMAIN	1042	1081	EGF-LIKE 26.		
FT	DOMAIN	1083	1122	EGF-LIKE 27.		
FT	DOMAIN	1126	1167	EGF-LIKE 28.		
FT	REPEAT	1168	1208	EGF-LIKE 29.		
FT	REPEAT	1209	1242	LIN/NOTCH 1.		
FT	REPEAT	1243	1282	LIN/NOTCH 2.		
FT	REPEAT	1282	1357	LIN/NOTCH 3.		
FT	REPEAT	1357	1428	ANK 1.		
FT	REPEAT	1428	1499	ANK 2.		
FT	REPEAT	1499	1570	ANK 3.		
FT	REPEAT	1570	1641	ANK 4.		
FT	REPEAT	1641	1712			
FT	REPEAT	1712	1783			

Alignment Scores:	
Pred. No.:	1.83
Score:	96.50
Percent Similarity:	31.46%
Best Local Similarity:	23.94%
Query Match:	8.29%
DB:	1
Length:	196
Matches:	51
Conservative:	16
Mismatches:	71
Indels:	75
Gaps:	13

US-09-809-545A-1_COPY_535_1143 (1-609) X NTC4_MOUSE (1-1964)

QY 57 TGTGGGCGCGGTTCACACCCCACTTTA----- 86
 Db 914 CywPrPrProGlyPheGlnGlyValLeuCyseGlnAspMetAspProGluProAsn 9333
 QY 87 ---TGCAGGACACGGGTCTGTGTGTGCCAGGACCAACAGAGGGATCTTCACATGACACTGG 143
 Db 934 ProCywHtHtAsGlySerThrCywValrProGlnProSerGly----- 947
 QY 144 CCCCACTTCACTGTATTACTTCTGCACATGCTCGCTTTCATATATGGGCGGCACTGC 2030
 Db 948 -----TyrValCyseGlnCywAlaProGlyTyrGlnGlyGlnAsn----- 960
 QY 204 TGCACCTGCATACCGAGGGGCGACCTTCGAGAGCGGTGTGCACACCGGTACAAACACTT 2636
 Db 961 CywSerTyrValLeuGlnAlaCyseGlnSerGlnProCywHtAsnHtAs----- 976

[illegible]

```
Search completed: March 12, 2003, 21:19:34
Job time : 22.5 secs
```



```

RP      SEQUENCE FROM N.A.
RA      Chen W., Winkelmann J.C.;
RT      "The Exon-Intron Organization of the Human HRNep1 Gene."
RL      Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF229057; AAL71904.1; -
DR      EMBL; AF109120; AAL83409.1; -
DR      EMBL; AF109107; AAL83409.1; JOINED.
DR      EMBL; AF109108; AAL83409.1; JOINED.
DR      EMBL; AF109109; AAL83409.1; JOINED.
DR      EMBL; AF109110; AAL83409.1; JOINED.
DR      EMBL; AF109111; AAL83409.1; JOINED.
DR      EMBL; AF109112; AAL83409.1; JOINED.
DR      EMBL; AF109113; AAL83409.1; JOINED.
DR      EMBL; AF109115; AAL83409.1; JOINED.
DR      EMBL; AF109116; AAL83409.1; JOINED.
DR      EMBL; AF109117; AAL83409.1; JOINED.
DR      EMBL; AF109118; AAL83409.1; JOINED.
DR      EMBL; AF109119; AAL83409.1; JOINED.
DR      InterPro; IPR000504; RNA_rec_mot.
DR      Pfam; PF000076; rrm; 1.
DR      SMART; SMO0360; RRM; 1.
DR      PROSITE; PS00030; RRM_1P; 1.
DR      PROSITE; PS00030; RRM_RNP_1; UNKNOWN.1.
SQ      SEQUENCE   418 AA;  44793 MW;  4706FCT40F75D01A CRC64;

```

Alignment scores:	
Pred. No.:	4,37e-88
Score:	1063.00
Percent Similarity:	98.52%
Best Local Similarity:	97.54%
Query Match:	91.32%
DB:	4
	Gaps: 0

[illegible]

Db	396	ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrArgGlyGlyTyrAsnArgPhe	415
Qy	601	GCTCCATAT	609
Db	416	AlaProTyr	418
RESULT 2			
QBR427			
ID	QBR427	PRELIMINARY;	PRT; 417 AA.
AC	QBR427;		
DT	01-JUN-2002	(TREMBLrel. 21, Created)	
DT	01-JUN-2002	(TREMBLrel. 21, Last sequence update)	
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)	
DE	Hexaribonucleotide binding protein 1.		
GN	HRNP1.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=SKELTAL MUSCLE;		
RA	Chen W., Winkelman J.C.;		
RL	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF191501; ALU83425.1;		
SQ	SEQUENCE 417 AA; 44638 MW; 96144A26BCDA888F CRC64;		

Alignment Scores:	
Pred. No.:	6,636-88
Score:	1061.00
Percent Similarity:	98.62%
Best Local Similarity:	97.54%
Query Match:	91.15%
DB:	11
	Gaps:
	Indels:
	0
	Gaps:
	417
	Matches:
	198
	Conservative:
	2
	Mismatches:
	3
	Indels:
	0
	Gaps:
	0

```

US-09-809-545A-1_COPY_535_1143 (1-609) x QBR427 (1-417)
QY      1  ATGACTAATAAAGAAGCGCGTGAACCCCTTACACCAATGGCTGGAAATTAAATCCAGTTGTG 60
Db      215 MetThrAnuLbYsThrValAsnProtyrThraSngLyTrpLbLeuAsnProValVal 234
QY      61  GCGCGGGCTCTACAGCCCGGACTTCTATAGCAGGCAACGGTGTCTGTGTGTGCCAGGCCAACAG 120
Db      235 GlyAlaValTYrSerProAspPheTYrLacLYthrValLeuLeuCYGlnAlaAsnGln 254
QY      121  GAGGGAATCTTCACAGTAAAGTGGCCCGACGTTCACTTGTATATACTTGTGCAATGCCCGC 180
Db      255 GluGlySerSerMetTYrSerGlyProSerSerLeuValTYrThrSerAlaMetProGly 274
QY      181  TTTCATATCCGGCGCGCGCATGCTGCAGCTGCATACCGAGGGGCTCACTTCAGGCGCT 240
Db      275 PheProTYrProAlaAlaSerAlaAlaAlaAlaTYrGlyAlaAlaIbLeuArgGlyArg 294
QY      241  GGTGCACCGTGTAACAACACCTTGAGAGCTGGCGGCCGCCACCCCAATCCGGGCTAT 300
Db      295 GlyArgThrGlyTYrAsnThrLeuArgAlaAlaAlaProProProProIleProAlaTYr 314
QY      301  GCGGAGTAGTGATCAAGAGCCACGTGTATGGCAATTAATTGCTACAGGCTGTACGCT 360
Db      315 GlyGlyValIValTYrGlnGlnIbProValTYrGlyAsnLysLeuLeuGlnGlyTYrAla 334
QY      361  GCATACCGCTACGCCAGCCCAACCCCTGCCCTCTCTGCTTACATGACATTAGGGA 420
Db      335 AlaTYrArgTYrAlaGlnIbProThrProAlaThrAlaAlaTYrSerAspSerTYrGly 354
QY      421  CGAGTTTAGTGGCGGACCCCTTACACACACACACTTGTCTCAGGCCCCACCTAGGCGGT 480
Db      355 ArgValTYrAlaAlaAspProtyrHisIbThrLeuAlaProAlaProIbTYrGlyVal 374
QY      481  GGTGCACGAATGCTTTTGGCCCTTGACCGATCCCAAGACTAGAGCACAATGCTGATGAT 540
Db      375 GlyAlaMetLeuAlaPheAlaProLeuThrAspAlaLYsThrArgSerHisAlaAsp 394

```

QY	541	GTGGGTCGCTTCCTTCCTTCATTCAGGCTGATTTATACAAAGGGGATACACCGTTT	601	GCTCCATAT	609	
Db	395	ValGlyLeuValIleuSerSerIleuGlnAlaSerIleTyraGlyGlyTyraAnaArgPhe	414			
QY	601	GCTCCATAT	609			
Db	415	AlaProTyx	417			
RESULT	3					
Q8TA3						
Q8TA3						
AC	Q8TA3					
DT	01-JUN-2002	(TREMBlrel. 21, Created)				
DT	01-JUN-2002	(TREMBlrel. 21, Last sequence update)				
DT	01-JUN-2002	(TREMBlrel. 21, Last annotation update)				
DE	Hexaribonucleotide binding protein 1 isoform alpha.					
GN	HRNBPI.					
OC	Homo sapiens (Human)					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.					
OK	NCBI_TaxID=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RA	Chen W., Chu Z.-L., Blough R.I., Liu L., Hoppes B., Winkelman J.C.;					
RT	"Molecular Cloning and Chromosomal Localization of a Novel Human					
RT	Brain, Heart and Skeletal Muscle Specific RNA Binding Protein Gene					
RT	Homologous to fox-1 in Caenorhabditis elegans."					
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RA	Chen W., Chu Z.-L., Blough R.I., Liu L., Hoppes B., Winkelman J.C.;					
RT	"Molecular Cloning and Chromosomal Localization of a Novel Human					
RT	Brain, Heart and Skeletal Muscle Specific RNA Binding Protein Gene					
RT	Homologous to fox-1 in Caenorhabditis elegans."					
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.					
RN	[3]					
RP	SEQUENCE FROM N.A.					
RA	Chen W., Winkelman J.C.;					
RT	"The Exon-Intron Organization of the Human HRNBPI Gene."					
RT	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.					
RL	EMBL; AF094849; AAL83405.1; -					
DR	EMBL; AF109120; AAL83408.1; JOINED.					
DR	EMBL; AF109107; AAL83408.1; JOINED.					
DR	EMBL; AF109108; AAL83408.1; JOINED.					
DR	EMBL; AF109109; AAL83408.1; JOINED.					
DR	EMBL; AF109110; AAL83408.1; JOINED.					
DR	EMBL; AF109111; AAL83408.1; JOINED.					
DR	EMBL; AF109112; AAL83408.1; JOINED.					
DR	EMBL; AF109113; AAL83408.1; JOINED.					
DR	EMBL; AF109114; AAL83408.1; JOINED.					
DR	EMBL; AF109115; AAL83408.1; JOINED.					
DR	EMBL; AF109116; AAL83408.1; JOINED.					
DR	EMBL; AF109117; AAL83408.1; JOINED.					
DR	EMBL; AF109119; AAL83408.1; JOINED.					
SD	SEQUENCE 392 AA; 42084 MW; 68F8CEBF2FC6D74 CRC64;					
Alignment Scores:						
Pred. No.:	2,21e-74	Length:	392			
Score:	912.00	Matches:	172			
Percent Similarity:	85.71%	Conservative:	2			
Best Local Similarity:	84.73%	Mismatches:	3			
Query Match:	78.35%	Indels:	26			
DB:	4	Gaps:	1			
US-09-809-545A-1_COPY_535_1143 (1-609) X Q8TA3 (1-392)						
QY	1	ATGACTAATAAAAGGCGGTGAACCCCTACACCAATGCGTGGAAATTAATCCAGTTGTG	60			
Db	216	MetTrAsnLysLysLysThValAsnProTyThrAsnGlyTrpLysLeuAsnProValVal	239			
QY	61	GGCGGCGCTTACAGCCCGGACTTCATATCAGGCAAGCGTGGTGTGCAAGCAACAG	120			
Db	236	GlyAlaValAlaTySerProIuPheTyraIaGlyThrValIleuLeuCySGlnAlaAsnGln	255			

OY		121	GAGGATATTCACGTACAGAGGGCCCGAGTTCACTTATAAATCTTGCAATGCCGTGC	180
Db		256	GlutylsersemetyIYrseAlaProSeSerleuValIYthrThrserAlametProGly	275
OY		151	TTTCATATCCGGCCGCCACTGTGACGTGCATACCAGAGGGCTCACCTTGAGGCCGT	240
Db		276	PheProTYrProAlaalaThnAlaAlaAlaAlaIeTYrArgGlyAlaHISLeuArGlyArg	295
OY		241	GGTGCACCCGTGAACAACACCTTAGAGCTGCCGGCCGCCACCCCACAAATCCGGCCTAT	300
Db		296	GlyAgtHmrValIYrasnthPhearGalalAlaLaProPropoPolleProAlaTYr	315
OY		301	GGCCGAGTAGTATCCAAGACCAGCTTAAGCCAAATAAATTGGCTACAGSGTGTTACGCT	360
Db		316	GlyGelValIvalIYrgInguInProValIYrcIYasnYslseuleuEngInGlyIYTyAla	335
OY		351	GCATACCCGCTACGCCCCAGCCACCCTTGCCACTGCTGCTGCTTACAGTACAAGTACGGA	420
Db		336	AlaTYrAgtTYrAlaGlnPrOthPrOAlathrAlaAlaIaIaTYrSerAspserTYrGly	355
OY		421	CGAGTTATNGCTGCGACCCCTTACCACACACACTTGCTCAGCCCCCACCCTACGGCGTT	480
Db		356	ArgValIYrAlaAlaAspProTYrhshIeAlaleuAlaProAlaProtnrTYrGlyVal	375
OY		481	GGTGCCATGAATGCTTTTGCCCCCTTGACCGAGTCCCAAAGACTAGAGAACCATGCTGATAT	540
Db		376	GlyAlaMet-----	378
OY		541	GTGGGTCTGCTTTCTTTCTTCAATGACAGGCTTAGTATATACCAAGGGGATACACCGTTTT	600
Db		379	-----AlaserIleTYrArgGlyIYrAsnarGphe	389
OY		601	GCTCCATAT 609	
Db		390	AlaProTYr 392	
RESULT 4				
OBTAF2				
ID	OBTAF2	PRELIMINARY;	PRT; 395 AA.	
AC	OBTAF2			
DT	01-JUN-2002 (TrEMBLrel. 21, Created)			
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)			
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)			
DE	Hemaphysoulonclotide binding protein 1 isoform beta.			
GN	HRNBP1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.			
OX	NCB1_TaxID=9606;			
RN	[1]			
RA	Chen W., Chu Z.-L., Blough R.I., Liu L., Hoppes B., Winkelmann J.C.;			
RT	"Molecular Cloning and Chromosomal Localization of a Novel Human			
RT	Brain, Heart and Skeletal Muscle Specific RNA Binding Protein Gene			
RT	Homologous to fox-1 in Caenorhabditis elegans."			
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Chen W., Winkelmann J.C.;			
RT	"The Exon-Intron Organization of the Human HRNBP1 Gene.";			
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF109106; AAL83406.1; -			
DR	EMBL; AF109120; AAL83407.1; JOINED			
DR	EMBL; AF109107; AAL83407.1; JOINED			
DR	EMBL; AF109108; AAL83407.1; JOINED			
DR	EMBL; AF109109; AAL83407.1; JOINED			
DR	EMBL; AF109110; AAL83407.1; JOINED			
DR	EMBL; AF109111; AAL83407.1; JOINED			
DR	EMBL; AF109112; AAL83407.1; JOINED			
DR	EMBL; AF109113; AAL83407.1; JOINED			
DR	EMBL; AF109114; AAL83407.1; JOINED			
DR	EMBL; AF109115; AAL83407.1; JOINED			

DR EMBL; AF109116; AAL83407.1; JOINED.
DR EMBL; AF109117; AAL83407.1; JOINED.
DR EMBL; AF109118; AAL83407.1; JOINED.
DR EMBL; AF109119; AAL83407.1; JOINED.
SQ SEQUENCE 395 AA; 42402 MW; 2259F9127AA1DE39 CRC64;

Alignment Scores:
Pred. No.: 7,566-66 Length: 395
Score: 818.00 Matches: 158
Percent Similarity: 87.85% Conservative: 1
Best Local Similarity: 87.29% Mismatches: 3
Query Match: 70.27% Indels: 19
DB: 4 Gaps: 1

US-09-809-545A-1_COPY_535_1143 (1-609) x 08TAF2 (1-395)

QY 1 ATGACTAATAAAGCGCGTGAACCCCTACCAATGGCTGGAATTAATCCAGTTGNG 60
DB 216 MettrAenLysLysThrValAsnProTyrThrAsnGlyTyrLysLeuAsnProValVal 235
QY 61 GCGCGGTCTACAGCCCGCACTTCTATGACAGCGAGCTGTTGTGCAGGCCAACAG 120
DB 226 GlyAlaValAllySerProGluPheTyrAlaGlyThrValLeuLeuCyGlnAlaAsnGln 255
QY 121 GAGGATCTTCATGATGACAGTGGCCCACTTCACTTGATATTAATCTTGCATGCTGCG 180
DB 256 GluGlySerSerMetTyrSerAlaProSerSerLeuValTyrThrSerAlaMetProGly 275
QY 181 TTTCATATCCGGCGCGCACTGCTGAGTGCATACCGAGGGGCTCACTTGAGGCGCGT 240
DB 276 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 295
QY 241 GGTCCACCGTGTACAAACACCTTCAGAGCTGGCGGCCCAACCCCAATCCCGGCTAT 300
DB 286 GlyAgtThrValTyrAsnThrPheArgAlaAlaAlaProProProProIleProAlaTyr 315
QY 301 GCGGAGTAGTGTATCAAGAGCCAGCTGATGCAATTAATCTTCAAGGTTGTTACGCT 360
DB 316 GlyGlyValValTyrGlnGlnProValTyrGlyAsnLysLeuLeuGlnGlyGlyTyrAla 335
QY 361 GCATACCGCTAGCCCGCAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 405
DB 336 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspArgAsnGln 355
QY 406 -----AGTACAGATTACGAGAGATT 426
DB 356 PheValPheValAlaAlaAspGluLysSerCysAsnThrSer-AlaValThrAspGluPh 375
QY 427 TATGCTGCCGACCCCTACCAACACACTTGTCTCCAGCCCCCACTTACGCGCTTGTGCC 486
DB 375 MetLeuProThrProThrThrThrHisLeuLeuGlnProProProThrAlaLeuValPr 395
QY 487 A 487
DB 395 o 395
RESULT 5
Q9SK10 PRELIMINARY: PRT: 376 AA.
AC Q9SK10. 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 40.6 kDa protein.
OS Macaca fascicularis (Crib eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Cercopitheciinae; Macaca.
OC Cercopitheciinae; Macaca.
OC NCB1_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TEMPORAL LOBE RIGHT;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,

RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB060859; BAB46877.1; -
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; Trm; 1.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
DR Hypothetical protein.
SQ SEQUENCE 376 AA; 40565 MW; DAA7A4D0D7B1030A CRC64;

Alignment Scores:
Pred. No.: 5,396-61 Length: 376
Score: 764.50 Matches: 150
Percent Similarity: 84.53% Conservative: 3
Best Local Similarity: 82.87% Mismatches: 8
Query Match: 65.68% Indels: 20
DB: 6 Gaps: 2

US-09-809-545A-1_COPY_535_1143 (1-609) x Q9SK10 (1-376)

QY 1 ATGACTAATAAAGCGCGTGAACCCCTACCAACATGGCTGGAATTAATCCAGTTGNG 60
DB 198 MettrAenLysLysThrValAsnProTyrThrAsnGlyTyrLysLeuAsnProValVal 217
QY 61 GCGCGGTCTACAGCCCGCACTTCTATGACAGCGAGCTGTTGTGCAGGCCAACAG 120
DB 218 GlyAlaValAllySerProGluPheTyrAlaGlyThrValLeuLeuCyGlnAlaAsnGln 237
QY 121 GAGGATCTTCATGATGACAGTGGCCCACTTCACTTGATATTAATCTTGCATGCTGCG 180
DB 238 GluGlySerSerMetTyrSerAlaProSerSerLeuValTyrThrSerAlaMetProGly 257
QY 181 TTTCATATCCGGCGCGCACTGCTGAGTGCATACCGAGGGGCTCACTTGAGGCGCGT 240
DB 258 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 277
QY 241 GGTCCACCGTGTACAAACACCTTCAGAGCTGGCGGCCCAACCCCAATCCCGGCTAT 300
DB 278 GlyAgtThrValTyrAsnThrPheArgAlaAlaAlaProProProProIleProAlaTyr 297
QY 301 GCGGAGTAGTGTATCAAGAGCCAGCTGATGCAATTAATCTTCAAGGTTGTTACGCT 360
DB 298 GlyGlyValValTyrGlnAspGlyPheTyrGlyAlaAsp---IleTyrGlyGlyTyrAla 316
QY 361 GCATACCGCTAGCCCGCAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 405
DB 317 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspArgAsnGln 336
QY 406 -----AGTACAGATTACGAGAGATT 426
DB 337 PheValPheValAlaAlaAspGluLysSerCysAsnThrSer-AlaValThrAspGluPh 356
QY 427 TATGCTGCCGACCCCTACCAACACACTTGTCTCCAGCCCCCACTTACGCGCTTGTGCC 486
DB 356 MetLeuProThrProThrThrThrHisLeuLeuGlnProProProThrAlaLeuValPr 376
QY 487 A 487
DB 376 o 376
RESULT 6
Q9UGW3 PRELIMINARY: PRT: 293 AA.
AC Q9UGW3. 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE D141P2.2 (Supported by GENSCAN) (Fragment).
GN RBM9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Whiteley M.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL049748; CAB63055.1; -.
DR HSSP; P11940; 1CVJ.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PSS0102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
FT NON_TER 1
FT SEQUENCE 293 AA; 30983 MW; F9A823DD70CC1004 CRC64;
US-09-809-545a-1_copy_535_1143 (1-609) x Q9UGW3 (1-293)
Alignment Scores:
Pred. No.: 5,94e-40 Length: 293
Score: 532.50 Matches: 114
Percent Similarity: 60.65% Conservative: 17
Best Local Similarity: 52.78% Mismatches: 44
Query Match: 45.75% Indels: 41
DB: Gaps: 7
QY 1 ATGACATATAAAGGCGCGTACACCAATGCGTGAATTAATCCAGTTGTG 60
Db 106 MetthrleuylslysmetvalThrProtyrAlaasnlytrpysleuserProvalVal 125
QY 61 GCGCGGCTACACGCCCGGACTTCTATGACGACGCGTGTGTCGACGCAAC- 117
Db 126 GlyAlaValTyrglyProgluLeuTyrrAlaAlaserPheGlnAlaAspValSerleu 145
QY 118 ---CAGAGGAGATCTTCATGATACAGTGGCCCGACTTACTT-----GTATATACTTCT 168
Db 146 GlyAsnspAlaAlaValProleuserGlyArgglylyleuThrTyrrleProleu 165
QY 169 GCAATGCGCTTTCATATCCG---GCCGCACCTGCTGACGATACCGAGGAGCT 225
Db 166 lIeIleProglPheProtyrProThrAlaAlaThrThrAlaAlaPheArgglyAla 185
QY 226 CACCTTGAGGCGGTGTCGACCGGTGTAACAACCTTCAGAGTGGCGGCCCGCCACC 285
Db 186 HistleuArglyArgglyArgThrValTyrglyAlaValArg---AlaValProProthr 204
QY 286 CCAATCCCGGCTTACGCGAGTGTATCAAGAGCCAGTGTATGCAATPAATTGCTA 345
Db 205 AlaIleProAlaTyrrProglylIleValleuGlnuProIleIleSerAlaIlePro 224
QY 346 CAGGTGGTACGCGTACGATACCGTACGCGCCAGCCCGCTGCGCACT-----CTT 393
Db 225 GlnIlylTyrrAlaAlaIleTyrrArgTyrrAlaGlnProAlaThrAlaThrAla 244
QY 394 -----GCTGCTCCTACAGTACAGTACAGGAGGAGTATGTCGCGACCC 441
Db 245 AlaAlaAlaAlaAlaAlaIleTyrrSerAspGlyTyrrGlyArgValTyrrAlaAspPro 264
QY 442 TACACACACACACTTGTCTCCGCCCCACCTACGCGCGTGTGTCATGATGCTTTGCG 501
Db 265 Tyr---HisAlaIleuAlaProAlaAlaSerTyrrGlyValGlyAlaVal----- 279
QY 502 CCTTGACCGATGCAAGACTAGAGGACATGCTATGATGCGGCTGCTTCTTCTCA 561
Db 279 ----- 279
QY 562 TTGACGCTAGTATATACAAAGGGGATACACCGTTTGTCTCATAT 609
Db 280 -----AlaserleuTyrrArgglyTyrrSerArgPheAlaProTyrr 293
RESULT 7
Q8WYB1 PRELIMINARY; PRT; 380 AA.

AC Q8WYB1;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hexaribonucleotide binding protein 2 (RNA binding motif protein 9).
GN HRNP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen W., Winkelman J.C.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=UTERUS;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF229058; AAL71905.1; -.
DR EMBL; BC025281; AAH25281.1; -.
DR InterPro; IPR000104; Antifreeze_1.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR PRINTS; PR00308; ANTIFREEZE1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PSS0102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
SQ SEQUENCE 380 AA; 40351 MW; 601FAE31E7C20715 CRC64;
US-09-809-545a-1_copy_535_1143 (1-609) x Q8WYB1 (1-380)
Alignment Scores:
Pred. No.: 2.43e-39 Length: 380
Score: 526.00 Matches: 116
Percent Similarity: 59.55% Conservative: 15
Best Local Similarity: 52.73% Mismatches: 43
Query Match: 45.19% Indels: 46
DB: Gaps: 8
QY 1 ATGACATATAAAGGCGGTGACCCCTTACACCAATGCGTGAATTAATCCAGTTGTG 60
Db 190 MetthrleuylslysmetvalThrProtyrAlaasnlytrpysleuserProvalVal 209
QY 61 GCGCGGCTACACGCCCGGACTTCTATGACGACGCGTGTGTCGACGCAAC- 117
Db 210 GlyAlaValTyrglyProgluLeuTyrrAlaAlaserPheGlnAlaAspValSerleu 229
QY 118 ---CAGAGGAGATCTTCATGATACAGTGGCCCGACTTCA-----CTT 156
Db 230 GlyAsnspAlaAlaValProleuserGlyArgglylyleuThrTyrrleProleu 249
QY 157 GTATATATCTTGCATATGCTGCTTTCATATCCG---GCCGCACCTGCTGACGTGCA 213
Db 250 lIeSerleuProleuValProglPheProtyrProThrAlaAlaThrThrAlaAla 269
QY 214 TACGAGGAGGCTCCTTCGAGCGCGGTGTCGACCGGTGTAACAACCTTCAGAGTGGCG 273
Db 270 PheArgglyAlaHisleuArgglyArgThrValTyrrGlyAlaValArg---Ala 288
QY 274 GCGCCCGCCACCCCAATCCCGGCTATGCGAGTATGATCAAGAGCCAGTATGCG 333
Db 289 ValProProthrAlaIleProAlaIleProglValValTyrrGlnAspGlyPheTyrrGly 308
QY 334 AATAAATGCTACAGGCGTGTACGTCATACCGCTACCGCCAGCCCGCTGCTCACT 393
Db 309 AlaAsp---leuTyrrGlyGlyTyrrAlaAlaIleTyrrArgTyrrAlaGlnProAlaThr 327
QY 394 -----GCTGCTCCTACAGTACAGTACAGGAGGAGTATGTCGCGACCC 429
Db 328 AlaAlaThrAlaAlaAlaAlaAlaIleTyrrSerAspGlyTyrrGlyArgValTyrr 347

QY 430 GCTGCCGACCCCTACACACACACTTGTCTCCAGCCCCCAGCTGAGGCTGTGTCATG 489
DB 348 ThrAlaAspProTyr---HisAlaLeuAlaProAlaAlaSerTyrGlyValGlyAlaVal 366
QY 490 AATGCTTTGGCCCTTGACCATGCCAAGACTAGAGCCATGCTGATGATGTGGTCTC 549
DB 366 ----- 366
QY 550 GTTCTTTCTTCATTCAGAGCTAGTATATACCAAGGGGATACACCGTTTGTCCATAT 609
DB 367 -----AlaSerLeuTyrArgGlyGlyTyrSerArgPheAlaProTyr 380
RESULT 8
Q8TD00 PRELIMINARY; PRT; 390 AA.
AC Q8TD00;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE RNA binding motif protein 9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OC NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21864496; PubMed=11875103;
RA Norris J.D., Fan D., Sheik A., McDonnell D.P.;
RT "A Negative Coregulator for the Human ER.;"
RL Mol. Endocrinol. 16:459-468(2002).
DR EMBL; AY072786; AL67150.1; -
SQ SEQUENCE 390 AA; 41374 MW; 40D6C8C40764B317 CRC64;

Alignment Scores:
Pred. No.: 2,446-39 Length: 390
Score: 526.00 Matches: 116
Percent Similarity: 59.55% Conservative: 15
Best Local Similarity: 52.73% Mismatches: 43
Query Match: 45.19% Indels: 46
Gaps: 8
US-09-809-545A-1_COPY_535_1143 (1-609) x Q8TD00 (1-390)

QY 1 ATGACTAATAAAAGCCGCTGAACCCCTACACCAATGCTGGAATTAATCCAGTTGTG 60
DB 200 MetTrnAsnLysLysMetValThrProTyrAlaAsnGlyTrpLysLeuSerProValVal 219
QY 61 GCGCGGCTTACAGCCCGCACTTCTTACAGGACCGCTGCTGTGTGCCAGGCCAAC--- 117
DB 220 GtAlaValTyrGlyProGluLeuTyrAlaAlaSerSerPheGlnAlaAspValSerLeu 239
QY 118 ---CAGGAGGAGTCTTCATGATGACAGTGGCCCACTCA-----CTT 156
DB 240 GtAlaAspAlaAlaValProLeuSerGlyArgGlyGlyIleAsnThrTyrIleProLeu 259
QY 157 GTATATACTTTCGCAATGCTGCTTTCATATCCG---GCCGCCACTGCTGACAGTGA 213
DB 260 IleSerLeuProLeuValProGlyPheProTyrProThrAlaAlaThrAlaAlaAla 279
QY 214 TACCGAGGGGCTACCTTGGAGGCGCTGTGCGACCGTGTACACACCTTCAGAGCTGG 273
DB 280 PheArgGlyAlaIleLeuAlaArgGlyArgGlyThrValTyrGlyIleAlaValArg---Ala 298
QY 274 GGGGCCCCCAGCCCAATCCCGGCTATGGCGAGTAGTATCAAGACAGCAGTATGGC 333
DB 299 ValProProThrAlaAlaIleProAlaTyrProGlyValValTyrGlnAspGlyPheTyrGly 318
QY 334 AATTAATCTACAGGCTGTTACGCTGATACCGCTACGCCAGCCAGCCCTGCTGCAC 393
DB 319 AlaAsp---LeuTyrGlyGlyTyrAlaAlaTyrArgTyrAlaGlnProAlaThrAlaThr 337
QY 394 -----GCTGCTGCTTACAGTACAGTTCAGGAGCAGTTTAT 429

DB 338 AlaAlaThrAlaAlaAlaAlaAlaAlaTyrSerAspGlyTyrGlyArgValTyr 357
QY 430 GCTGCCGACCCCTACACACACACTTGTCTCCAGCCCCCAGCTGAGGCTGTGTCATG 489
DB 358 ThrAlaAspProTyr---HisAlaLeuAlaProAlaAlaSerTyrGlyValGlyAlaVal 376
QY 490 AATGCTTTGGCCCTTGACCATGCCAAGACTAGAGCCATGCTGATGATGTGGTCTC 549
DB 376 ----- 376
QY 550 GTTCTTTCTTCATTCAGAGCTAGTATATACCAAGGGGATACACCGTTTGTCCATAT 609
DB 377 -----AlaSerLeuTyrArgGlyGlyTyrSerArgPheAlaProTyr 390
RESULT 9
Q923W8 PRELIMINARY; PRT; 377 AA.
AC Q923W8;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative RNA-binding protein fxh.
GN FXH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21294761; PubMed=11401487;
RA Lieberman A.P., Friedlich D.L., Harrison G., Howell B.W., Jordan C.L.,
RA Breddlove S.W., Fischbeck K.H.;
RT "Androgens Regulate the Mammalian Homologues of Invertebrate Sex
RT Determination Genes tra-2 and fox-1."
RL Biochem. Biophys. Res. Commun. 282:499-506(2001).
DR EMBL; AF387332; AA64287.1; -
DR MGD; MGI:1935973; Fxn.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; xrm; 1.
DR PROSITE; PS0102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN 1.
SQ SEQUENCE 377 AA; 40165 MW; 249FA5BA6A0D155A CRC64;

Alignment Scores:
Pred. No.: 2,996-39 Length: 377
Score: 525.00 Matches: 116
Percent Similarity: 60.19% Conservative: 14
Best Local Similarity: 53.70% Mismatches: 44
Query Match: 45.10% Indels: 42
Gaps: 8
US-09-809-545A-1_COPY_535_1143 (1-609) x Q923W8 (1-377)

QY 1 ATGACTAATAAAAGCCGCTGAACCCCTACACCAATGCTGGAATTAATCCAGTTGTG 60
DB 191 MetTrnAsnLysLysMetValThrProTyrAlaAsnGlyTrpLysLeuSerProValVal 210
QY 61 GCGCGGCTTACAGCCCGCACTTCTTATGACAGGACCGTGTGTGCCAGGCCAAC--- 117
DB 211 GtAlaValTyrGlyProGluLeuTyrAlaAlaSerSerPheGlnAlaAspValSerLeu 230
QY 118 ---CAGGAGGAGTCTTCATGATGACAGTGGCCCACTTCACTT-----GTATATACTTCT 168
DB 231 GtAlaAsnAlaAlaValProLeuSerGlyArgGlyGlyIleAsnThrTyrIleProLeu 250
QY 169 GCATATGCTGCTTTCATATCCG---GCCGCCACTGCTGACAGTACAGGAGGCT 225
DB 251 IleIleProGlyPheProTyrProThrAlaAlaThrAlaAlaAlaPheArgGlyAla 270
QY 226 CACCTTCAGAGCGCTGTGTCAGCGTGTACACACCTTCAGAGCTGCGGCGCCACCC 285
DB 271 HisLeuArgGlyArgGlyArgThrValTyrGlyAlaValArg---AlaValProProThr 289


```
QY 286 CCAATCCCGGCTATGCGGAGTAGTATCAAGACCGATGATGCAATTAATTGCTA 345
DB 290 AAlaIleProAlaTyPrroGlyValValTyrgIlnsPglPheTyrgIAlaasp---Leu 308
QY 346 CAGGTGTGTTACGCTGATFACCGCTAGCCGACCCCTGCGACT----- 393
DB 309 TyrgIlyGlyTyraIaAlaTyraTyrgIlnsPglPheTyrgIAlaAlaThrAla 328
QY 394 -----GCTGCTGCTTACAGTATGACGAGGAGTTTATGCTGCCGACCC 441
DB 329 AAlaIaAlaAlaAlaAlaAlaAlaTyrsersPglTyrgIAlaValTyThAlaaspPro 348
QY 442 TACACACACACACTGCTCCAGCCGCCACCTAGCGCGTGGTGCATGAATGCTTTGGC 501
DB 349 Tyr---HisAlaIeAlaProAlaAlaSerTyrgIValGlyAlaVal----- 363
QY 502 CCTTGACCGATGCCAGACTAGAGCCATGATGATGCGGTCTGCTTTCTTCA 561
DB 363 ----- 363
QY 562 TTGACGCTAGTATATACAGGGGAGATACACCGTTTGTCTCATAT 609
DB 364 -----AlaSerLeuTyraGlyGlyTyrsersAlaPheAlaProTyr 377

RESULT 10
Q8V162 PRELIMINARY; PRT; 377 AA.
AC Q8V162:
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hexaribonucleotide binding protein 2.
GN FXH OR HRNBP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RA SEQUENCE FROM N.A.
RA Chen W., Winkelman J.C.;
RA Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF229055; AAL71902.1; -.
DR MGD; MGI:1933973; Fxh.
DR InterPro; IPR00104; Antifreeze 1.
DR InterPro; IPR00504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR PRINTS; PR00308; ANTIFREEZE2I.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PSS0102; RRM; 1.
DR PROSITE; PSS0030; RRM_RNP_1; UNKNOWN 1.
SQ SEQUENCE 377 AA; 40183 MW; 249E72CAF47C505A CRC64;

Alignment Scores:
Pred. No.: 2.99e-39 Length: 377
Score: 525.00 Matches: 116
Percent Similarity: 60.19% Conservative: 14
Best Local Similarity: 53.70% Mismatches: 44
Query Match: 45.10% Indels: 42
Gaps: 8
DB: 11

US-09-809-545a-1_copy_535_1143 (1-609) x Q8V162 (1-377)
QY 1 ATGACTATAAAGAGCGCTGACCCCTACACCAATGCTGGAATTAATCCAGTTGTG 60
DB 191 MethranlsylsMetValThrProTyraIaasnGlyTPrlylsleuSerProValVal 210
QY 61 GCGCGGTCTACAGCCCGCACTTATGACGACGAGTGTGTCGACGACCAAC--- 117
DB 211 GlyIaValValTyrgIlnsPglPheTyraIaAlaSerPheGlnAlaAspValSerIeu 230
QY 118 ---CAGGAGGATCTTCATGATGACGCGCCCACTTACTT-----GTAATAATCTTCT 168
DB 231 GlyaasnGlnAlaAlaValProMetserGlyTyrgIlyIleasnThrTyrlasProIeu 250
```

```
QY 169 GCATGCTGCTTCCATATCCG---GCCGCACTGCTGCAAGTGCATACGAGGAGCT 225
DB 251 IleIleProGlyPheProTyPrroThrAlaAlaThrAlaAlaAlaPheAlaGlyAla 270
QY 226 CACCTTGAGCGCGTGTGCGACCGGTATACACCTTCAGAGCTGGCGGCCCAACC 285
DB 271 HisIeAlaTyrgIlnsPglPheTyraIaAlaValTyrgIAlaValAlaProProThr 289
QY 286 CCAATCCCGGCTATGCGGAGTAGTATCAAGACCGATGATGCAATTAATTGCTA 345
DB 290 AAlaIleProAlaTyPrroGlyValValTyrgIlnsPglPheTyrgIAlaasp---Leu 308
QY 346 CAGGTGTGTTACGCTGATFACCGCTAGCCGACCCCTGCGACT----- 393
DB 309 TyrgIlyGlyTyraIaAlaTyraTyrgIlnsPglPheTyrgIAlaAlaThrAla 328
QY 394 -----GCTGCTGCTTACAGTATGACGAGGAGTTTATGCTGCCGACCC 441
DB 329 AAlaIaAlaAlaAlaAlaAlaAlaTyrsersPglTyrgIAlaValTyThAlaaspPro 348
QY 442 TACACACACACTGCTCCAGCCGCCACCTAGCGCGTGGTGCATGAATGCTTTGGC 501
DB 349 Tyr---HisAlaIeAlaProAlaAlaSerTyrgIValGlyAlaVal----- 363
QY 502 CCTTGACCGATGCCAGACTAGAGCCATGATGATGCGGTCTGCTTTCTTCA 561
DB 363 ----- 363
QY 562 TTGACGCTAGTATATACAGGGGAGATACACCGTTTGTCTCATAT 609
DB 364 -----AlaSerLeuTyraGlyGlyTyrsersAlaPheAlaProTyr 377

RESULT 11
Q96D26 PRELIMINARY; PRT; 358 AA.
AC Q96D26:
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Similar to fox-1 homolog (C. elegans).
GN Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RA SEQUENCE FROM N.A.
RA Straussberg R.;
RA TISSUE=ADRENAL GLAND;
RA Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013115; AAL13115.1; -.
DR InterPro; IPR00504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR PROSITE; PSS0102; RRM; 1.
DR PROSITE; PSS0030; RRM_RNP_1; UNKNOWN 1.
SQ SEQUENCE 358 AA; 37907 MW; BA4B3740C178E441 CRC64;

Alignment Scores:
Pred. No.: 5.54e-39 Length: 358
Score: 522.00 Matches: 115
Percent Similarity: 60.19% Conservative: 15
Best Local Similarity: 53.24% Mismatches: 44
Query Match: 44.85% Indels: 42
Gaps: 8
DB: 4

US-09-809-545a-1_copy_535_1143 (1-609) x Q96D26 (1-358)
QY 1 ATGACTATAAAGAGCGCTGACCCCTACACCAATGCTGGAATTAATCCAGTTGTG 60
DB 172 MethranlsylsMetValThrProTyraIaasnGlyTPrlylsleuSerProValVal 191
QY 61 GCGCGGTCTACAGCCCGCACTTATGACGACGAGTGTGTCGACGACCAAC--- 117
DB 192 GlyIaValValTyrgIlnsPglPheTyraIaAlaSerPheGlnAlaAspValSerIeu 211
```

Qy	118	-----CAGGAGGGATCTTCATGATGACAGTGGCCCGCACTTACTT-----GTAATATACTTCT	169
Db	212	glyaenapalaalaaalavalproleuserdgiyarglgiyilieaenthrtyrileproleu	231
Qy	169	GCATATGCTGGCTTTCATATTCG---GCCGCCATCTGTCGAGTGCATACCGAGGAGCT	225
Db	232	llellepproglpneprotyrprothrhlalathrhlalalalalalalalalalalphearglyala	255
Qy	226	CACCTTGCAGGCGGTGTGCGACCGGTGTAACAACACCTTCAGAGCTGCGGCGCCCGCCACCC	285
Db	252	hlsleuafargllyarggllyargthrvaltyrgllyalavalarg---Alavalproprothr	270
Qy	286	CCAATCCCGGGCTATGGGGAGTACTGTATCAAGAGCCAGTGTACCAATAAATGTGTA	345
Db	271	Alallepproalatyrrpoglyvalvaltyrghlnspsglpnyethryrgllyalalasp---Leu	285
Qy	346	CAGGTGTTCACGTGATACCGCTACGCGCCAGCCACCGCTGCCACGTCT-----	396
Db	290	tyrglgiylyrlyal	305
Qy	397	-----GCTGCTTACAGTGAAGTATACGAGACGATTATGCTGCGGACCCC	444
Db	310	Ala	325
Qy	442	TACCAACACACACTGTGCTTCAGCGCCCAACCTACGAGCGTGTGGTGCATCAATGCTTTTGGC	501
Db	330	tyr---hlsleuafargllyarggllyargthrvaltyrghlnspsglpnyethryrgllyalalasp	344
Qy	502	CCCTTGACCGATGCCAGACTAGAGGACCATGCTGATGTGGGCTTCGTTCTTTCTTCA	561
Db	344	-----	344
Qy	562	TTGCAGCTGATGATATACCAAGCGGGATACACCGTTCGCTCCATAT	609
Db	345	-----AlaSerLeuTyraArgIyArgIyIySerArgPheAlaProTyrr	358
RESULT 12			
096NL7			
ID	096NL7	PRELIMINARY;	PRT; 450 AA.
AC	096NL7		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	CDNA FJ336651 file, clone D3OST20010184, highly similar to RNA binding		
DE	motif protein 9.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,		
RA	Futuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,		
RA	Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,		
RA	Okutsu T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,		
RA	Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,		
RA	Matsumoto K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,		
RA	Kawachi K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,		
RA	Suzuki Y., Sugano S., Negahari K., Masuno Y., Nagai K., Isogai T.,		
RT	"NEBO human cDNA sequencing project."		
RL	Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.		
DR	EMBL; AK055213; BAB70875.1; --		
DR	InterPro: IPR00504; RNA_rec_mot.		
DR	Pfam: PF00076; rrm; 1.		
DR	PROSITE: PS50102; RRM; 1.		
DR	PROSITE: PS00030; RRM_RNP_1; UNKNOWN_1.		
SO	SEQUENCE 450 AA; 47326 MW; 173D30C55D5CD762 CRC64;		
Alignment Scores:			
Pred. No.:	1,08e-38	Length:	450
Percent Similarity:	519.00	Matches:	115
	58.64%	Conservative:	14

Best Local Similarity:	52.27%	Mismatches:	45
Query Match:	44.59%	Indels:	46
DB:	4	Gaps:	8
US-09-809-545A-1_COPY_535_1143 (1-609) x 096NL7 (1-450)			
QY	1	ATGACTATAAAGGCGGTGAACCCCTACACCAATAGCGGAAATTAATTCAGTTGTG	60
DB	260	MechtrianLysLysMetValThrProLysAlaAsnLysLysLysSerProValAla	279
QY	61	GGCGGCTCTACACCCCGAATTCTATGACGAGCACGGTGTCTGTGCCAGGCGAAC	117
DB	280	GlyAlaValTyrGlyProGluLeuTyrAlaAlaLysSerPheGlnAlaSerLeu	299
QY	118	--CAGAGGAGTCTTCATGATACAGTGGCCCCAGTTCA-----CTT	156
DB	300	GlyAsnAlaAlaAlaValProLeuSerGlyArgGlyGlyLeuAsnThrTyrIleProLeu	319
QY	157	GNATTAATTCTTGCAATGCGGTTCCTCATATCCG--GCGGCACTGTCCAGCGCA	213
DB	320	IleSerLeuProLeuValProGlyPheProTyrProThrAlaAlaThrAlaAlaAla	339
QY	214	TACCGAGGGGCTCACTTTCGAGCGCGTGTGCGACCGTGTACAACAACCTTCAGAGTCCG	273
DB	340	PheArgGlyAlaHisLeuArgGlyArgGlyArgThrAlaTyrGlyAlaValArg--Ala	358
QY	274	GGCGCCCAACCCCAATCCGCGCTATGGGGAGTATGATCAAGAGCCAGTGTATGCC	333
DB	359	ValProProThrAlaIleProAlaTyrProGlyValAlaTyrGlnAspGlyPheTyrGly	378
QY	334	AATAATTGCTACAGGGGTGACGTGACATCCGCTACGCGCCAGCCACCCCTGCGACT	393
DB	379	AlaAsp--LeuTyrGlyGlyTyrAlaAlaTyrArgTyrAlaGlnProAlaThrAlaThr	397
QY	394	-----GCTGCTGCTTACAGTACGATGACGATTAACGACGAGTTTAT	429
DB	398	AlaAlaThrAlaAlaAlaAlaAlaAlaAlaAlaTyrGlyAspGlyTyrGlyArgValTyr	417
QY	430	GCTGCCAGCCCTTACACACACACACTTGTCTCCAGCCCGACCTTACGGCGTTGCTGCATG	489
DB	418	ThrAlaAspProTyr--HisAlaLeuAlaProAlaAlaSerTyrGlyAlaGlyAlaVal	436
QY	490	AATGTTTTGGCGCTTGACCGATGCCAAGCTACGAGCCATGCTGATGATGGGCTTC	549
DB	436	-----	436
QY	550	GTTCTTTCTCATTCAGAGCTAGTATATACCAAGGGGATACAACGTTTGTCTCATAT	609
DB	437	-----AlaSerLeuTyrArgGlyGlyTyrSerArgPheAlaProTyr	450
RESULT 13			
Q8VI61 PRELIMINARY; PRT; 303 AA.			
AC	Q8VI61:		
DT	01-MAR-2002 (TREMBLrel. 20, Created)		
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Hexaribonucleotide binding protein 3 (Fragment).		
GN	HRNBP3.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
OX	NCBI_TaxId=10090;		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BALB/C;		
RL	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF229056; AAL71903.1; -		
DR	InterPro; IPR00504; RNA_rec_mot.		
DR	Pfam; PF00076; rrm; 1.		
DR	SMART; SM00360; RRM; 1.		
DR	PROSITE; PS50102; RRM; 1.		

DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
 FT NON_TER 1 1
 SQ SEQUENCE 303 AA; 32864 MW; 74B0AF9F87E935 CRC64;

Alignment Scores:
 Pred. No.: 5.27e-33 Length: 303
 Score: 456.00 Matches: 95
 Percent Similarity: 48.33% Conservative: 6
 Best Local Similarity: 45.45% Mismatches: 32
 Query Match: 39.18% Indels: 76
 DB: 11 Gaps: 4

US-09-809-545A-1_COPY_535_1143 (1-609) x Q8V161 (1-303)

QY 7 AATAAAGGCGCTGAACCCCTACACCAATGCTGAAATTAATTCAGTTGTGGCGCG 66
 Db 123 AsnAsnAlaThrAlaArgValMetThrAsnGlyTrrpysLeuAsnProValValGlyThr 142

QY 67 GTCTACAGCCCGCACTTCTATGACAGCAGCGGTGCTGTGCGACGACGAGGGA 126
 Db 143 ValTyrGlyProGluPheTyrAlaValThr----- 152

QY 127 TCTTCATGTACAGTGGCCCGCACTTCTATATACTTCTGCAATGCTGCTTTCCA 186
 Db 153 -----SerPhePro 155

QY 187 TATCCGCGCCGCACTGCTGACGTACACGAGGCGCTGACCTTGCAGCGCGTGTGCG 246
 Db 156 TyrProThrThrGlyThrAlaValAlaTyrArgGlyThrshLeuArgGlyAlaArg 175

QY 247 ACCGTGTAACAACACCTTCAGAGCTGCGCGCGCCCGCAATCCGCGCTATGGCGGA 306
 Db 176 AlaValTyrAsnThrPheArgAlaAlaProProProPheProThrTyrGlyAla 195

QY 306 ----- 306

Db 196 AlaLeuGluGlnThrLeuValLysMetProValProTrrAlaGlyLeuAlaProCysPro 215

QY 306 ----- 306

Db 216 LeuProProGlnGlnThrProGluProAlaTyrProThrSerProAlaPheProLeu 235

QY 307 -----GTAAGTATACAGACGACGATGATGGCAATTAATTGCTA 345
 Db 236 SerCysProPheAlaSerArgValValTyrGlnAspGlyPheTyrGlyAlaGlu---1le 254

QY 346 CAGGTGTTACGCTGCATACCGCTAGCGCCACCGCCCTGCACTGCTGCTGCTAC 405
 Db 255 TyrGlyGlyTyrAlaAlaTyrArgTyrAlaGlnProAlaAlaAlaAlaAlaTyr 274

QY 406 AGTACAGTTACGACGAGTTAT---GCTGCGACCGCTACCAACACCACTGCTGCCA 462
 Db 275 SerAspSerTyrGlyArgValTyrAlaAlaAspProTyrshshThrIleGlyPro 294

QY 463 GCCCCACCTACAGCGCTGTCATG 489
 Db 295 ThrAlaThrTyrSerIleGlyThrMet 303

RESULT 14
 Q8TCMO PRELIMINARY; PRT; 263 AA.

AC Q8TCMO; 08TCMO; 21, Created)
 DT 01-JUN-2002 (T-EMBLrel. 21, last sequence update)
 DT 01-JUN-2002 (T-EMBLrel. 21, last sequence update)
 DE Hypothetical 28.6 kDa protein (Fragment).
 GN DKF2547L059.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1] SEQUENCE FROM N.A.

RC TISSUE=BRAIN;
 RA Bloeker H., Boecker M., Brandt P., Mewes H.W., Weil B., Wiemann S.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL713700; CAD28499.1; -
 KW Hypothetical protein

FT NON_TER 263 263
 SQ SEQUENCE 263 AA; 28565 MW; 56C2D853846BCE1 CRC64;

Alignment Scores:
 Pred. No.: 1.75e-23 Length: 263
 Score: 351.00 Matches: 65
 Percent Similarity: 97.06% Conservative: 1
 Best Local Similarity: 95.59% Mismatches: 2
 Query Match: 30.15% Indels: 0
 DB: 4 Gaps: 0

US-09-809-545A-1_COPY_535_1143 (1-609) x Q8R2T5 (1-263)

QY 1 ATGACTAATAAAGGCGCTGAACCCCTACACCAATGCTGAAATTAATTCAGTTGTG 60
 Db 196 MetThrAsnLysThrAlaAsnProTyrThrAsnGlyTrrpysLeuAsnProValVal 215

QY 61 GCGCGGCTTACAGCCCGCACTTCTATGACAGCAGCGGTGCTGTGCGACGCAACAG 120
 Db 216 GlyAlaValTyrSerProGluPheTyrAlaGlyThrValLeuGlyGlnAlaAsnGln 235

QY 121 GAGGATCTTCATGTACAGTGGCCCGCACTTCTATATACTTCTGCAATGCTGCG 180
 Db 236 GluGlySerSerMetCysTyrSerAlaProSerSerLeuValTyrThrSerAlaMetProGly 255

QY 181 TTTTCATATCCGCGCGCACTGCT 204
 Db 256 PheProTyrProAlaAlaThrAla 263

RESULT 15
 Q8R2T5 PRELIMINARY; PRT; 435 AA.

AC Q8R2T5; 21, Created)
 DT 01-JUN-2002 (T-EMBLrel. 21, last sequence update)
 DT 01-JUN-2002 (T-EMBLrel. 21, last sequence update)
 DE Similar to RNA binding motif protein 9.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1] SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC027263; AAH27263.1; -
 SQ SEQUENCE 435 AA; 46252 MW; 62A355605FF8A5B4 CRC64;

Alignment Scores:
 Pred. No.: 1.01e-18 Length: 435
 Score: 299.00 Matches: 80
 Percent Similarity: 53.94% Conservative: 9
 Best Local Similarity: 48.48% Mismatches: 55
 Query Match: 25.69% Indels: 21
 DB: 11 Gaps: 7

US-09-809-545A-1_COPY_535_1143 (1-609) x Q8R2T5 (1-435)

QY 1 ATGACTAATAAAGGCGCTGAACCCCTACACCAATGCTGAAATTAATTCAGTTGTG 60
 Db 259 MetThrAsnLysThrAlaAsnProTyrThrAsnGlyTrrpysLeuSerProValVal 278

QY 61 GCGCGGCTTACAGCCCGCACTTCTATGACAGCAGCGGTGCTGTGCGACGCAAC---117
 Db 279 GlyAlaValTyrGlyProGluLeuTyrAlaAlaSererPheGlnAlaAspValSerLeu 298

QY 118 ---CAGGAGGATCTTCATGTACAGTGGCCCGCACTTCTATATACTTCT 168
 Db 118 ---GTAATATCTTCT 168

```

Db 299 GlyAenGIuAlaAlaValProLeuSerGIyArgGIyIleAsnThrTyrlleProLeu 318
Qy 169 GCAATGCTGCTGCTTCATATCCG--GCCGGCACTGCTGCAGCTGCATACGAGGGGCT 225
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 319 IleIleProGIyPheProTyPProThrAlaIleThrAlaIleAlaIlePheArgGIyAla 338
Qy 226 CACCTTGAGGCGCTGTGCGCACCGGTGACAAACACCTTCAGAGCTGCCGCGCCACCC 285
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 339 HisLeuArgGIyArgGIyArgThrValTyrgIyAlaValArg--AlaValProProThr 357
Qy 286 CCAATCCCGGCTATGGGGAGTAGTATCAAGAGCCAGTGTATGGCAATAAATTGCTA 345
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 AlaIleProAlaTyPProGIyAlaIlePheMetGln--ProThrAspMetHisSerIleu 376
Qy 346 CAGGGTGTTACGCTGCATACCGCTACGCCAGCCACCCCTGCCACTGCTGCTGCT-- 403
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 377 -----LeuGlnProGlnPro-GlnLeuLeuGlnProle 387
Qy 404 -----ACAGTGACAGTTACGAGACGAGTTTATGCTGCCGACCCTACGACACACACA 453
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 387 uGlnProLeuThrAlaThrValThrAlaGIyCystrGlnLeuThrProThrMetProSe 407
Qy 454 CTGGCTCCAGCCC 466
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 407 rProLeuProPro 411

```

Search completed: March 12, 2003, 22:23:32
 Job time : 43.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 12, 2003, 21:18:34 ; Search time 13.5 Seconds
(without alignments)
2654.601 Million cell updates/sec

Title: US-09-809-545A-1_COPY_535_1143
Perfect score: 1164
Sequence: 1 atgacataaataaaagcgctt.....acacacgtttgtccatcat 609

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues
Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p model -DEV=xjh
-Q=/cgn2_1/USFTO_spool/US090945/rnat_11032003_130947_12341/app_query.fasta_1.775
-DB=Issued Patents AA -QFMT=fastan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=0.0 -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USFR=US090945.0CEN_1.13 @rnat_11032003_130947_12341 -NCPU=6 -ICPU=3
-NO_XLTX -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.dep.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.dep.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.dep.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.dep.*
5: /cgn2_6/ptodata/1/iaa/PCUS.COMB.dep.*
6: /cgn2_6/ptodata/1/iaa/backflist.dep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	624	53.6	330	4	US-09-145-391-2
2	315.5	27.1	366	3	US-09-176-657-2
3	100.5	8.6	349	4	US-09-343-011B-1
4	96	8.2	760	1	US-08-195-152-2
5	95.5	8.2	1172	1	US-08-313-288B-19
6	94.5	8.1	1964	4	US-09-467-997-1
7	93	8.0	1444	4	US-09-252-292C-27
8	92.5	7.9	362	1	US-08-437-027-21
9	92.5	7.9	365	1	US-08-437-027-20
10	92.5	7.9	656	2	US-08-343-443B-2
11	92.5	7.9	656	4	US-09-214-564A-4
12	92	7.9	219	4	US-09-527-345-2

13	91	7.8	346	4	US-09-343-011B-2	Sequence 2, Appli
14	87.5	7.5	980	4	US-09-442-100-8	Sequence 8, Appli
15	87	7.5	379	1	US-08-552-142A-11	Sequence 11, Appli
16	87	7.5	1706	2	US-08-459-568-2	Sequence 2, Appli
17	87	7.5	1706	2	US-08-399-411-2	Sequence 2, Appli
18	87	7.5	1706	3	US-08-516-859A-2	Sequence 2, Appli
19	87	7.5	1706	4	US-09-586-472-2	Sequence 2, Appli
20	87	7.5	1706	4	US-09-528-706-2	Sequence 2, Appli
21	86.5	7.4	262	3	US-08-946-914-14	Sequence 14, Appli
22	86.5	7.4	262	3	US-09-656-450-14	Sequence 14, Appli
23	86.5	7.4	1274	4	US-09-095-443-2	Sequence 2, Appli
24	86.5	7.4	4302	3	US-08-658-136-5	Sequence 5, Appli
25	86.5	7.4	4302	4	US-09-052-469-8	Sequence 8, Appli
26	86.5	7.4	4339	4	US-09-052-469-6	Sequence 6, Appli
27	86	7.4	207	4	US-09-336-536-50	Sequence 50, Appli
28	86	7.4	245	4	US-09-336-536-48	Sequence 48, Appli
29	86	7.4	462	3	US-08-875-944B-5	Sequence 5, Appli
30	86	7.4	462	4	US-09-116-049-4	Sequence 4, Appli
31	86	7.4	469	4	US-08-313-288B-15	Sequence 15, Appli
32	85.5	7.3	572	6	5256770-7	Patent No. 5256770
33	85.5	7.3	1185	4	US-09-041-886-23	Sequence 23, Appli
34	85	7.3	275	1	US-08-312-870-7	Sequence 7, Appli
35	85	7.3	275	1	US-08-431-387-3	Sequence 3, Appli
36	85	7.3	275	1	US-08-322-677A-7	Sequence 7, Appli
37	85	7.3	275	1	US-08-322-676-7	Sequence 7, Appli
38	85	7.3	275	1	US-08-460-343B-74	Sequence 74, Appli
39	85	7.3	275	1	US-08-398-028B-74	Sequence 74, Appli
40	85	7.3	275	2	US-08-504-265B-90	Sequence 90, Appli
41	85	7.3	275	2	US-08-140-083A-9	Sequence 9, Appli
42	85	7.3	275	2	US-08-865-203-8	Sequence 8, Appli
43	85	7.3	275	2	US-09-135-658-3	Sequence 3, Appli
44	85	7.3	275	2	US-07-849-420-8	Sequence 8, Appli
45	85	7.3	275	3	US-08-898-218-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-145-391-2
; Sequence 2, Application US/09145391
; Patent No. 6194171
; GENERAL INFORMATION:
; APPLICANT: Puist, Stefan M.
; APPLICANT: Shibata, Hiroki
; TITLE OR INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
; TITLE OR INVENTION: Products Related Thereto and Methods of Using Same
; FILE REFERENCE: CE 3093
; CURRENT APPLICATION NUMBER: US/09/145,391
; CURRENT FILING DATE: 1998-09-01
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-145-391-2

Alignment Scores:
Pred. No.: 1.6e-54
Score: 624.00
Percent Similarity: 89.78%
Best Local Similarity: 86.6%
Query Match: 53.61%
DB: 4
Gaps: 0
Length: 330
Matches: 119
Conservative: 4
Mismatch: 12
Indels: 2

US-09-809-545A-1_COPY_535_1143 (1-609) x US-09-145-391-2 (1-330)
OY 1 ATGACTATAAAGCGCGTGAACCCCTACACCAATGCTGAATTAATTCAGTTGTG 60
DB 196 MetThrAsnIlyslsYstHrValAsnProIyThrAsnIlyTrpIlyslsLeuAsnProValVal 215
OY 61 GGCAGCGTCTACAGCCCCGACCTTATGACGACGCGTCTGTGTGCGACGCCAACAG 120

Db	216	GLYALAVALTYRSEPRGGLUPHERYALAELYTHRALVALLEUCYSGINALAENGIN	235
Qy	121	GAGGATATTCACGATTAAGAGGCCACGATCACTGTATATATCTTGCAATGCCGCG	180
Db	236	GLUGLSEETSEMEYTRSERALAPROSEETSERUEVALYTRNRSERALAKEPROGLY	255
Qy	181	TTTCCATATCCGGCCGCCACTGCTGTGCAGCTGCATACCGAGGGGCTACCTTGAGGCGGT	240
Db	256	PHENOTYRPROALALATHRALALALALATRYRGLYALANISLEUAARGLYARG	275
Qy	241	GATGCACCGTGTCACACCTTCAGAGCTGGGGGGCCCCACCCCAATCCGGGCGCTAT	300
Db	276	GLYAGTNHVALTYRASNTHPRHEATGALALALAPROFORPROFOLLEPROALATYR	295
Qy	301	GGCGGAGTAGTGTATCAAGAGCCAGTGTATGCGAATAAATTGCTACAGGGTGTACGCT	350
Db	296	GLYGLYVALIATYRPRGGLYTPRIELEUTRIPCYSARGHISLEU--TRTPRIEUCYSYC	315
Qy	361	GCATACCGGCTACGGCCAGCCACCCCTGGCCACTGTGCTGCTACAG	407
Db	315	YELLEPTOLEUADRPROALATYRPROCYANHSCYBAQCYBLEGIN	330

```

1 RESULT 2
2 US-09-176-657-2
3 Sequence 2, Application US/09176657
4 Patent No. 6020164
5 GENERAL INFORMATION:
6 APPLICANT: Bandman, Olga
7 APPLICANT: Tang, Y. Tom
8 APPLICANT: Corley, Neil C.
9 APPLICANT: Guegler, Karl U.
10 APPLICANT: Lu, Alina
11 APPLICANT: Baughn, Mariah R.
12 TITLE OF INVENTION: HUMAN RNA BINDING PROTEINS
13 FILE REFERENCE: PF-0611 US
14 CURRENT APPLICATION NUMBER: US/09/176,657
15 CURRENT FILING DATE: 1998-10-21
16 NUMBER OF SEQ ID NOS: 9
17 SOFTWARE: PERL Program
18 SEQ ID NO 2
19 LENGTH: 366
20 TYPE: PR7
21 ORGANISM: Homo sapiens
22 FEATURE: -
23 OTHER INFORMATION: 1250374
24 US-09-176-657-2

```

Alignment Scores:	
Pred. No.:	1,41e-23
Score:	315.50
Percent Similarity:	49.05%
Best Local Similarity:	42.86%
Query Match:	27.10%
DB:	3
Length:	366
Matches:	90
Conservative:	13
Mismatches:	59
Indels:	48
Gaps:	10

US-09-809-545A-1_COPY_535_1143 (1-609) X US-09-176-657-2 (1-366)

Qy	1	ATGACTATAAAAAAGCCGCTAACCCCTCAACCAATGGCTGGAAATTAAATTCAGTTGTG	60
		
Db	190	MecthranLnLbVbMecValThPrOtyrAlaIngLyTrpLbLseuSePrOvalVal	209
Qy	61	GGCGCGCTCAACAGCCCGCATTTCTATGACAGACAGCGGTCTGTGTCCAGGCCAAC--	117
		
Db	210	GIyAlaValTyrgLyProGluLeuTyrrAlaIseTerSerPheGlnAlaApValSerLeu	229
Qy	118	--CAGGAGGATCTTCATGTACAGTGGCCCCAGTTCACTT-----GTATATCTTCT	168
		
Db	230	GIyAenApAlaAlaValProLseuSerGIyrgLyGIyLeuAenThTyrrLeProLseu	249
Qy	169	GCAAGCGCTGCTTCCATATCCG---CGCCGCACTGCTGACGTGATACCGAGGGCT	225
		
Db	250	IleIleProGlyPheProTyrrProThrAlaIethrThrrAlaAlaAlaPheAerGgLyAla	269

Oy	226	CACCTTGCAGGCCCGCTGTGCGACCGGTAAACAACCTTCAGAGCTGGCGGGCCCAACC	285
Db	270	Hlsneuartrgclglaagclglaagthnvaltylclglaalarg---	AlaValProProthr 288
Oy	286	CCAAATCCCGGCGCTATGGCGGAGTAGTGTATCAAGAGCCAGCTGTATGGCAATTAATTGCTA	345
Db	289	AlaIleProAlaIyrProgllyValaIspMetGln---	ProThrAspMetHisSerLeuLeu 307
Oy	346	CAGGCTGTTACGCTGCATACCGCTACGCCACGCCACCCCTGGCATTTGGTCTGCTCT--	403
Db	308	-----LeuGlnProIlePro--ProleuLeuGlnProLe	318
Oy	404	-----ACAGTGCAGCTTACGAGACGAGATTATGCTCCGACCCCTCAACACACACA	453
Db	318	uGlnProleuthrValthrValthrValmetalaIaglyCySthGlnProthrProthr-----	355
Oy	454	CTTGCTCCAGCCCCCACACTACCGCGGTGGTGCCATGAATGCTTTGGCGCCCTTGACCGAT	513
Db	336	-----MetProleuProLe	340
Oy	514	GCCAGACTAGGAACCATGCTGATGATGTGGGTCTGCTTTCTTCTTCATTGACAGGCTAGT	573
Db	340	uPro---LeuValMetGluLeuValaLeuTrp-----	ArgVal 351
Oy	574	ATATACCAAGGGGATACAAACCGTTTG	601
Db	351	IlyrthGlnValaIaIaThrIaIaIspLeu	360

```

RESULT 33
US-09-343-011B-1
Sequence 1, Application US/09343011B
Patent No. 6300473
GENERAL INFORMATION:
Applicant: Stephane Richard
TITLE OF INVENTION: SLM-1 AND SLM-2; NOVEL
FILE REFERENCE: A32561
CURRENT APPLICATION NUMBER: US/09/343,011B
CURRENT FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: CA 2265271
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Fastseq for Windows Version 3.0
SEQ ID NO. 1
LENGTH: 349
TYPE: PRT
ORGANISM: Mus musculus
US-09-343-011B-1

```

Pred. No.:	0.0501	length:	34
Score:	100.50	Matches:	56
Percent Similarity:	35.92%	Conservative:	18
Best Local Similarity:	27.18%	Mismatches:	72
Query Match:	8.63%	Indels:	
DB:		Gaps:	9
			4

US-09-809-545A-1_COPY_535_1143 (1-609) X US-09-343-011B-1 (1-349)

```

QY 48 AATATCAATTGTGGGCGCGCTACACCCCGAATTCTA----- 86
      ::::|
Db 183 GluSerGlyArgGlyArgGlyIleArgGlyValGlyIleThrProThrAlaPro 202
      ::::|
QY 87 -----TGCAGGACGGTGCCTGTGTGCCAGGACCAACGAGGAGATCTTCATGTACAG 140
      |||
Db 203 SerArgGlyArgGlyGlyAlaValProProProProPro-ProGlyArgGlyValLeuTh 222
      |||
QY 141 TGGCCCCAGTTCACTGTATATATCTTGCATACCT-----GGCTT 182
      ::::|
Db 222 rProArgGlyIleThrValThrArgGlyAlaLeuProValProIleAlaArgGlyVala 242
      ::::|
QY 183 TGCATATCCGGCGCCGACCTGCTGCAGCTGCATACCGAGGGGCTCACCTTGAGAGCCGTGG 242
      |||

```

Db 242 lProthPrroarglAaarglyThrAlaValProgly----- 255
QY 243 TCCGACCCGTGACAACTTCAGAGCTGCGCGCCGCCAACCCCAATCCCGCTATGG 302
Db 256 -----TyrArg---AlaProProProProAlaHisAspAlaTyrG1 268
QY 303 CGGAGTAGTATACAGAGCCAGTGTATGCGAATAATTGCTACAGGGGTGATCGCTGC 362
Db 268 uG1uTyG1TyTyrAspAspGly---TyrGlyGlyGlyTyrAspAspGlnThrTyrG1uA1 287
QY 363 ATACCGGTAGCCGACCCGACCTGCGACGCTGCTCCACAGTACAGTACATTCAGACG 422
Db 287 ATYrAspAspSerTyrValThrProThrGlnSerValProGluTyrTyrAspTyrGlyH1 307
QY 423 AGTTATGCTGCGCAGCCCTACACACACACTGTCTCAGCCGCCACCTACGGCGTTGG 482
Db 307 sGlyValAsnGlnAspAlaTyr----- 314
QY 483 TGCCATGATGCTTTTGCGCCCTTGACCGATGCCAAGTACAGACCTGCTATGATGT 542
Db 315 -----AspSerTyrAlaProGluGluTyrAlaThrThrArg----- 326
QY 543 GGGCTGCTCTTCTTCATGTCAGGCT-----AGTATATACCAAGGGGATACAA 593
Db 327 -----SerSerLeuysAlaProProProArgSerAlaArgGlyGlyTyr-- 341
QY 594 CCGTTTGTCTCATAT 609
Db 342 -ArgGluHisProTyr 346
RESULT 4
US-08-195-152-2
Sequence 2, Application US/08195152
Patent No. 5679541
GENERAL INFORMATION:
APPLICANT: Bonini, Nancy M.
APPLICANT: Leiserson, William M.
TITLE OF INVENTION: PROGRAMMED CELL DEATH ANTAGONIST
TITLE OF INVENTION: PROTEINS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobach, Teet, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
City: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/195,152
FILING DATE: 14-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Treacartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-59551/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 760 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-195-152-2

Alignment Scores:
Pred. No.: 0.182 Length: 760
Score: 96.00 Matches: 42
Percent Similarity: 37.29% Conservative: 24
Best Local Similarity: 23.73% Mismatches: 61
Query Match: 8.25% Indels: 50
DB: 1 Gaps: 6
US-09-809-545a-1_copy_535_1143 (1-609) x US-08-195-152-2 (1-760)
QY 124 GGATCTTCATATGAC-----ACTGGCCCACTTCACTT 156
Db 223 G1ySerAsnLeuTyrGlyCySerSerAlaSerAsnProLeuAspGlyAlaValAla 242
QY 157 GTATATCTTCTGCAATACCTGCTTTCATATCCGCGCCGCTGCTGACGTGATAC 216
Db 243 ValAsnSerSerAlaVal-----AlaAlaAlaAlaAlaValTyr 256
QY 217 CGAGGGCTACCTTCGAGGCGGTGTGCGACCGGTACACACCTTCAGACTGCGGCG 276
Db 257 AspGlyLysHis-----AspTyrTyrTyrTyrAsnSerMetGlnGlnTyrThr 272
QY 277 CCCCCACCCCAATCCCGGCTATGCGGAGTAGTATACAGACAGCTATGCGCAT 336
Db 273 ProProProPheTyrSerGlyTyrGlyThrProTyrAlaAlaAlaThrAlaAlaArgGln 292
QY 337 AAATTGCTACAGGGGTGTTAGCTGATACCGCTACCGCCGACCCCTGACACTGCT 396
Db 293 AlAlaSerMetGluProGlyAlaAlaAlaAlaAlaAlaAlaTyrLeuThrProSerTyrAla 312
QY 397 GCTGCC----- 402
Db 313 AlaSerGlyAsnAsnSerGlnLeuTyrSerSerProTyrAlaGlyTyrAsnAsnPhe 332
QY 403 -----TACAGTACAGTATACGACAGCTTATGCTGCCGAC 438
Db 333 G1yGlnGlnAspTyrGlyGlyTyrTyrAsnGlnGlnTyrGlyAsnTyrTyrSerProAla 352
QY 439 CCGTACACACACACTGCTCCAGCCCGC-----ACCTAGCGGCTGCT-- 483
Db 353 AsnTyrSerProTyrAlaValAsnSerProSerSerSerAlaSerHisGlyHisGlyPhe 372
QY 484 GCCATGATGCTTTTGCGCCCTTGACCGATGCCAAGCTACAGACCATGCT 534
Db 373 HisValAlaAlaSerSerAsnLeuSerGluSerProThrAspThrHisSer 389
RESULT 5
US-08-313-288B-19
Sequence 19, Application US/08313288B
Patent No. 5750502
GENERAL INFORMATION:
APPLICANT: Jessell, Thomas M. and Avihu Klar
TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
City: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/313,288B
FILING DATE: January 5, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.

GENERAL INFORMATION:
APPLICANT: Boudec, Phillipe
APPLICANT: Rodgers, Matthew
APPLICANT: Dumas, Florence
TITLE OF INVENTION: Mutated hydroxyphenylpyruvate dioxygenase. DNA
TITLE OF INVENTION: sequence and isolation of plants which contain such a
FILE REFERENCE: 5500*31
CURRENT APPLICATION NUMBER: US/09/252,292C
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 08/982,772
PRIORITY FILING DATE: 1997-12-02
NUMBER OF SEQ ID NOS: 31
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 27
LENGTH: 444
TYPE: PRT
ORGANISM: Zea mays
US-09-252-292C-27

Alignment Scores:
Pred. No.: 0.306 Length: 444
Score: 93.00 Matches: 51
Percent Similarity: 32.59% Conservative: 22
Best Local Similarity: 22.77% Mismatches: 69
Query Match: 7.99% Indels: 82
DB: 4 Gaps: 10

US-09-809-545A-1_COPY_535_1143 (1-609) x US-09-252-292C-27 (1-444)

QY 163 ACTTCGCAATGCTGGCTTCATATCCGCGCCACTGCTGCGATGACGAGG 222
DB 112 ThrAlaAlaLeuProSerPheSerAlaAlaAlaAlaArgPheAlaAlaSerGly 131
QY 223 GCTCAGCTTGGAGG-----CGTGTGCGACCGCTGTACAACACTTCAGAGCTGCG 273
DB 132 LeuAlaValAlaArgAlaValAlaAlaSerAlaValAlaAspAlaPheArgAlaSer 151
QY 274 GCGCCCGCCACCCCAATCCCGGCTATGCGCGAGTA-----GTG 312
DB 152 ValAlaAlaGlyAlaArgProAlaPheGlyProValAspLeuGlyArgGlyPheArgLeu 171
QY 313 TATCAAGAGCGCATGTATGCAATTAATTCCTACAGGAGTGTACGCTGATACCGCTAC 372
DB 172 AlaGlyValGlyLeuLeuTyrglyAspValValLeuArg-----TyrValSerTyr----- 187
QY 373 GCCCAGCCCAACCCCTGCGACTGCTGCTGCTACAGTACGACGATTAACGACGATT----- 426
DB 188 -----ProAspGlyAlaAlaGlyGluProPheLeuProGlyPheGlyValAlaSer 205
QY 427 -----TATGCTGCGAGCCCTACACGAC-----ACA 453
DB 206 ProGlyAlaAlaAspTyrglyLeuSerArgPheAspHisIleValGlyAsnValProGlu 225
QY 454 CTTCGCTCAGCCCGCCACTAC----- 474
DB 226 LeuAlaProAlaAlaAlaAlaTyrglyPheAlaGlyPheThrGlyPheHisGluPheAlaGluPhe 245
QY 475 -----GCGCTGTGCTGCATG----- 489
DB 246 ThrThrGluAspValGlyThrAlaGluSerGlyLeuAsnSerMetValLeuAlaAsn 265
QY 490 -----AATGCTTTGGCGCTTCGACCGAT-----GCCAAGACTAGAGCGCAT 531
DB 266 SerGluAsnValLeuLeuProLeuAsnGluProValHisGlyThrTyrsArgSerGln 285
QY 532 -----GCT 534
DB 286 IleGlnThrPheLeuAspHisIleGlyGlyProGlyValGlnHisMetAlaLeuAlaSer 305
QY 535 GATGATGTGGGTCTGCTTTCTTTCATTCGAGGCTAGTATATACCAAGGGGATCAAC 594
DB 306 AspAspValLeuArgThrLeuArgGluMetGlnAlaArgSerAlaMetGlyGlyPheGlu 325

QY 595 CGTTTGTCTCA 606
DB 326 PheMetAlaPro 329

RESULT 8
US-08-437-027-21
Sequence 21, Application US/08437027
Patent No. 5670317
GENERAL INFORMATION:
APPLICANT: Landanyi, Marc
TITLE OF INVENTION: A DIAGNOSTIC TEST FOR TEST FOR THE DESMOPLASTIC
TITLE OF INVENTION: SMALL ROUND CELL TUMOR
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/437,027
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 46416/JPM/CCA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 362 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-437-027-21

Alignment Scores:
Pred. No.: 0.322 Length: 362
Score: 92.50 Matches: 48
Percent Similarity: 33.72% Conservative: 10
Best Local Similarity: 27.91% Mismatches: 73
Query Match: 7.95% Indels: 41
DB: 1 Gaps: 9

US-09-809-545A-1_COPY_535_1143 (1-609) x US-08-437-027-21 (1-362)

QY 85 TATGACGACGCGTGTGTGCGAGCC---AACCAGAGGATTTTCATGATACAGT 141
DB 29 TyrAlaGlnThr-----GlnAlaTyrglyGlnGlnSerTyrglyThrTyrgly 45
QY 142 GGGCCCGATTCATGTATATCTTCGCAATCGCTGCTTCATATCCGGCGCCACT 201
DB 46 GlnProThrAspValSerTyrglnAlaGlnThrThrAlaThrTyrglyGlnThrAla 65
QY 202 GCTCAGCTGCGATACGAGGGGCTCACTCGAGCGCGTGTGCGACCGTACACACC 261
DB 66 TyrAlaThrSerTyr-----GlyGlnProProThrGlyTyThrThr 79
QY 262 TTCAGAGCTCGGCGCCCGCCCAATCCGCGCTATGCGGAGTACGTATCAAGAG 321
DB 80 -----ProThrAlaProGln-----AlaTyrsGln 88


```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/06123
; FILING DATE: 20-MAY-1992
; ATTORNEY/AGENT INFORMATION:
;   NAME: Weiser, Gerard J.
;   REGISTRATION NUMBER: 19,763
;   REFERENCE/DOCKET NUMBER: 989,6121P
;   TELEPHONE: 215-875-8383
;   TELEFAX: 215-875-8394
;   INFORMATION FOR SEQ ID NO: 2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 656 amino acids
;       TYPE: amino acid
;       TOPOLOGY: linear
;   MOLECULE TYPE: protein
;
US-08-343-443B-2

Alignment Scores:
Pred. No.: 0.39 Length: 656
Score: 92.50 Matches: 48
Percent Similarity: 33.72% Conservative: 10
Best Local Similarity: 27.91% Mismatches: 73
Query Match: 7.95% Indels: 41
DB: Gaps: 9

US-09-809-545A-1_COPY_535_1143 (1-609) x US-08-343-443B-2 (1-656)
QY 85 TATGACGAGCAGCGCTGTGTGCGAGCC---AACCAGAGGAGATCTTCATATGACAGT 141
   ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 29 TYRALAGLINTHR-----GlnAlaTyrglynglnseryrglythrygly 45
QY 142 GGGCCGACCTGATATATCTTGTGCAATGCGCTTCCATATCCGGCGCCACT 201
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 46 GlnProthrspylserythrglnalaglnthrrhralathryrglynglnthrala 65
QY 202 GCTGACCTCATACCGAGGGGCTCACCTTCGAGCGCGTGTGCGACCGTGTACACACC 261
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 66 Tyralathrseryr-----glylnpropothrglythrythr 79
QY 262 TTCAGAGCTGGCGGCCCCCAATCCGGCTATGGCGGAGTAGTATCAAGAG 321
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 80 -----ProthralaProgln-----AlaTySergln 88
QY 322 CCAGTGATGCAATAATGCTACAGGCGTACGCTGATACCGCTACGCCAGCC 381
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 89 Provalglnlytyr-----glythrglyalatyraaspthrthrrhrala 103
QY 382 ACCCTGCACTGCTGCTGCTACAGTACAGAGTTCAGAGGAGTTATGCTGCCAGCCC 441
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 104 ThrValthrthrrhrglnala-----SeryrrAlaAlaGlnSer 116
QY 442 TACCACCACTGCTGCTCCAGCCCCCACTACGGCGTGTGTCGATGAATGCTTTGCG 501
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 117 AlaTyrglythrglnProalaTyrrProalaTyrglyln-----Gln 130
QY 502 CCCTGACCGATGCGAGGAGCCATGATGATGAGGCTGCTGCTCTTTCTTCA 561
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 131 ProAlaAlaThrAlaProthrrArgProGlnAspGlyasnlySProthrgluThrSergln 150
QY 562 TTGCAGGCTAGTATATCAAGGGGAGTACACCGT 597
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 151 ProGlnSer-----ThrglyGlyTyrrasnGln 160

RESULT 11
US-09-214-564A-4
; Sequence 4, Application US/09214564A
; Patent No. 6150515
; GENERAL INFORMATION:
; APPLICANT: Sharp, Phillip A.
; APPLICANT: Zhou, Qiang
; TITLE OF INVENTION: TAT-SF. Cofactor For Stimulation Of Transcriptional
; FILE REFERENCE: Elongation By HIV-1 TAT
```

```
; FILE REFERENCE: M0656/7042
; CURRENT APPLICATION NUMBER: US/09/214,564A
; CURRENT FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: US 60/021,218
; PRIOR FILING DATE: 1996-07-03
; PRIOR APPLICATION NUMBER: US 60/033,152
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: PCT/US97/11713
; PRIOR FILING DATE: 1997-07-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 656
; TYPE: PRT
; ORGANISM: Homo sapiens
;
US-09-214-564A-4

Alignment Scores:
Pred. No.: 0.39 Length: 656
Score: 92.50 Matches: 48
Percent Similarity: 33.72% Conservative: 10
Best Local Similarity: 27.91% Mismatches: 73
Query Match: 7.95% Indels: 41
DB: Gaps: 9

US-09-809-545A-1_COPY_535_1143 (1-609) x US-09-214-564A-4 (1-656)
QY 85 TATGACGAGCAGCGCTGTGTGCGAGCC---AACCAGAGGAGATCTTCATATGACAGT 141
   ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 29 TYRALAGLINTHR-----GlnAlaTyrglynglnseryrglythrygly 45
QY 142 GGGCCGACCTGATATATCTTGTGCAATGCGCTTCCATATCCGGCGCCACT 201
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 46 GlnProthrspylserythrglnalaglnthrrhralathryrglynglnthrala 65
QY 202 GCTGACCTCATACCGAGGGGCTCACCTTCGAGCGCGTGTGCGACCGTGTACACACC 261
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 66 Tyralathrseryr-----glylnpropothrglythrythr 79
QY 262 TTCAGAGCTGGCGGCCCCCAATCCGGCTATGGCGGAGTAGTATCAAGAG 321
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 80 -----ProthralaProgln-----AlaTySergln 88
QY 322 CCAGTGATGCAATAATGCTACAGGCGTGTACGCTGATACCGCTACGCCAGCC 381
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 89 Provalglnlytyr-----glythrglyalatyraaspthrthrrhrala 103
QY 382 ACCCTGCACTGCTGCTGCTACAGTACAGAGTTCAGAGGAGTTATGCTGCCAGCCC 441
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 104 ThrValthrthrrhrglnala-----SeryrrAlaAlaGlnSer 116
QY 442 TACCACCACTGCTGCTCCAGCCCCCACTACGGCGTGTGTCGATGAATGCTTTGCG 501
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 117 AlaTyrglythrglnProalaTyrrProalaTyrglyln-----Gln 130
QY 502 CCCTGACCGATGCGAGGAGCCATGATGATGAGGCTGCTGCTCTTTCTTCA 561
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 131 ProAlaAlaThrAlaProthrrArgProGlnAspGlyasnlySProthrgluThrSergln 150
QY 562 TTGCAGGCTAGTATATCAAGGGGAGTACACCGT 597
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 151 ProGlnSer-----ThrglyGlyTyrrasnGln 160

RESULT 12
US-09-527-345-2
; Sequence 2, Application US/09527345
; Patent No. 6331413
; GENERAL INFORMATION:
; APPLICANT: Shepherd, Paul O.
; APPLICANT: Adler, David A.
; TITLE OF INVENTION: SECRETED SALIVARY ZS163 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/527,345
```

;; CURRENT FILING DATE: 1999-03-17
;; PRIOR APPLICATION NUMBER: US 60/124,820
;; PRIOR FILING DATE: 1999-03-17
;; NUMBER OF SEQ ID NOS: 9
;; SOFTWARE: FASTSEQ for Windows Version 3.0
;; SEQ ID NO 2
;; LENGTH: 219
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-527-345-2

Alignment Scores:
Pred. No.: 0.307 Length: 219
Score: 92.00 Matches: 53
Percent Similarity: 33.91% Conservative: 6
Best Local Similarity: 30.46% Mismatches: 56
Query Match: 7.90% Indels: 59
DB: Gaps: 10

US-09-809-545a-1_copy_535_1143 (1-609) x US-09-527-345-2 (1-219)
QY 24 CCCCTACACCAATGGCTGGAATTAATCAAGTTGGCGGCTCTACAGCCGCACTT 83
DB 33 ProlenHisProSerLeuAenIleProTyrGlyIle---ArgAenIleProProleu 51
QY 84 CTATGACGACGAGTGTCTGTGTCGCCAGCCAAACGAGAGGAGATCTTCATGATACAGTGG 143
DB 52 TyrTyrArgProValAenThrValProSerTyrProGlyAenThrTyrThrAspThrG1 71
QY 144 CCCGATTCACCTGTATATCTTCTGCAATGCTGCTTTCATATCGGCGCCGCACTGC 203
DB 71 YleuProSerTyrProTyrPileLeuThrSerProGlyPheProTyr----- 86
QY 204 TGCAGCTGCATACCGAGGGGCTCACTTCGAGGC----- 237
DB 87 -----ValTyr-----HisIleArgGlyPheProleuA1aThrGlnLeuAenVa 101
QY 238 -----CGTGTGTCGACCGCTGAC-----AACAGCTTCGAGCTGC 272
DB 101 LProProleuProProA1aGlyPheProPheValProProSerArgPheSerAlaA1 121
QY 273 GCGCGCCCA-----CCCCCAATCCCGCTATAGCGGAGTAGTGTATCAAGACCTACT 326
DB 121 AAlaIaProA1aIaIaProIleA1aIa-----GluProA1 134
QY 327 GTATGCGCAATAATTGCTACAGGGGTGTTAGCTGATACCGCTACGCGCACCCACC 386
DB 134 A-----AlaA1aIaProleuth 140
QY 387 TGCCACTGCTGCTGCTACAGTACGAGTTACGAGAGATTATGCTGGAGCCCTACCA 446
DB 140 rAlaIaThrProValA1aIaIaGluProA1aIaGlyA1aProValA1aIaIaGluPro----- 158
QY 447 CCACACACTGCTCCAGCCCACTCACTGCGCTGCTGCC 486
DB 159 -----AlaA1aGluA1aPro-----ValGlyA1a 166

RESULT 13
US-09-343-011B-2
; Sequence 2, Application US/09343011B
; Patent No. 6300473
; GENERAL INFORMATION:
; APPLICANT: Stephane Richard
; TITLE OF INVENTION: SLM-1 AND SLM-2: NOVEL
; FILE REFERENCE: SAME6-LIKE MAMMALIAN PROTEINS
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US/09/343,011B
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2

;; LENGTH: 346
;; TYPE: PRT
;; ORGANISM: Mus musculus
US-09-343-011B-2

Alignment Scores:
Pred. No.: 0.448 Length: 346
Score: 91.00 Matches: 32
Percent Similarity: 41.67% Conservative: 8
Best Local Similarity: 33.33% Mismatches: 46
Query Match: 7.82% Indels: 10
DB: Gaps: 4

US-09-809-545a-1_copy_535_1143 (1-609) x US-09-343-011B-2 (1-346)
QY 178 GCGTTTCATATCCGCGCCGCTATGCACTGCACTACGAGGGGCTCACTTGGAGGC 237
DB 221 GlyThrProThrProA1aGlyValLeuSerThr-----ArgGlyProValSerArgGly 238
QY 238 CGTGT-----CGCACCGTGTACAAACNC---TTCAGAGCTGCGGCG 276
DB 239 ArgGlyLeuLeuThrProArgAlaArgIyValProProThrGlyTyrArgProProPro 258
QY 277 CCCCACCCCAATCCCGGCTATGCGGAGTAGTATCAAGAGCCAGTATGCAAT 336
DB 259 ProProProThrGlnGluThrTyrGlyGlyTyrAspTyrAspAspGly---TyrGlyThr 277
QY 337 AAATGTCACAGGGTGTACGCTGATACCGCTACGCGCCAGCCACCCCTGCACTGCT 396
DB 278 AlaTyrAspGluGlnSerTyrAspSerTyrAspAspSerTyrSerThrProA1aGlnSer 297
QY 397 GCTGCTACAGTACAGATTACGAGACGAGTTATGCTGCCGACCCCTAC 444
DB 298 AlaAlaAspTyrTyrAspTyrGlyHisGlyLeuSerGluAspAlaTyr 313

RESULT 14
US-09-442-100-8
; Sequence 8, Application US/09442100
; Patent No. 6359193
; GENERAL INFORMATION:
; APPLICANT: Xu, Tian
; APPLICANT: Tao, Wufan
; APPLICANT: Wang, Weiyl
; APPLICANT: Zhang, Sheng
; APPLICANT: Yu, Wan
; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF LATS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/442,100
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/411,111
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6523-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090

```

; TELEFAX: (212) 869-9741/8864
; TELETYPE: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 980 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-09-442-100-8

Alignment Scores:
Pred. No.: 1,41 Length: 980
Score: 87.50 Matches: 50
Percent Similarity: 36.04% Conservative: 21
Best Local Similarity: 25.38% Mismatches: 81
Query Match: 7.52% Indels: 45
DB: Gaps: 9

US-09-809-545A-1_COPY_535_1143 (1-609) x US-09-442-100-8 (1-980)
QY 76 CCCGACTTCTATGACGAGCGGTCGTCGTCGACGAGCGCAACGAGGATCTTCATG 135
DB 108 Proserhegluglythnglylalauleuproserthynglnleuglygylalasn 127
QY 136 TACAGTGGCCCGACGTCATGTAATATCTTCCATGCTT-----GCCTTT 183
DB 128 Tyrcyluglyproalalaleu-----glugluecProAryGlnTyrcleuAspHe 144
QY 184 CCAATGCGGCGCGGCGACGTCGTCGTCATACGAGGCGGCTTCATCGAGCGCGGT 243
DB 145 leuAheProgly-----AlaGlyAlaGlyThrsiglyalaglnalanhsglnhePro 162
QY 244 CGCAGCGGTGTAACAACACCTTCGAGCTGGCGGCGCGCGCGCGCGCGCGCGCG 294
DB 163 Prolysglytyrserthr-----AlaValGluProseralanhserProGlyThrhIs 180
QY 295 --GCCATGCGGAGAGTAGTGTATCAAGACCGAGTGTATGCG----- 333
DB 181 TyrclyarglyhlsleuLeusergluglnserglytyrglyValGlnArgseSer 200
QY 334 -----ATAAATGCTACAGAGGTGTTAGCGCTGATACCGCTACCGCTACCG 378
DB 201 PheGlnasnlystrProProAspAlatyserSerMetAlaLysAlaGlnGlyPro 220
QY 379 -----CCACCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 426
DB 221 ProAlaSerleuthrPheProAlaHisAlaGlyLeuTyrrhAlaSerhHisHlyse 240
QY 427 TATGCTCCGACCGCTTACACACACACACTTGTCTCCAGCGCGCGCGCGCGCG 477
DB 241 AlaAlaThrProProGlyAlaHisProleuHisValLeuGlyThrhArgGlyProThrPhe 260
QY 478 GTTGTGTCATGATGCTTTTGGCGCGCTGACCGCATGCCAAGACTAGAGCGCAT 537
DB 261 ThrGlyIuserSerAlaGlnAlaValLeuAlaProSerArg----- 274
QY 538 GATGTGGGTCTCGCTTCTTCTTCTTCAATGCGAGGTATATACCAAGGGGA 588
DB 275 -----AasnSerleuasnAlaAspLeuTyrglyLeuGly 285

RESULT 15
US-08-552-142A-11
; Sequence 11, Application US/08552142A
; Patent No. 5695995
; GENERAL INFORMATION:
; APPLICANT: Weintraub, Harold M.
; APPLICANT: Lee, Jacqueline E.
; APPLICANT: Tapscott, Stephen J.
; APPLICANT: Hollenberg, Stanley J.
; TITLE OF INVENTION: Neurogenic Differentiation (Neurod) Genes
; TITLE OF INVENTION: and Proteins
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Christensen O'Connor Johnson KindnessP.L.L.C
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/552,142A
; FILING DATE: 02-NOV-1995
; CLASSIFICATION: 514
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,238
; FILING DATE: 06-MAY-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US95/05741
; FILING DATE: 08-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Broderick, Thomas F.
; REGISTRATION NUMBER: 31,332
; REFERENCE/DOCKET NUMBER: FPCR-1-8933
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100
; TELEFAX: 206-225-0709
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 379 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-552-142A-11

Alignment Scores:
Pred. No.: 1,16 Length: 379
Score: 87.00 Matches: 38
Percent Similarity: 33.52% Conservative: 23
Best Local Similarity: 20.88% Mismatches: 73
Query Match: 7.47% Indels: 48
DB: Gaps: 6

US-09-809-545A-1_COPY_535_1143 (1-609) x US-08-552-142A-11 (1-379)
QY 40 TCGAATTAATATCAAGTTGTGGCGCGGCTGTACAGCGCGCATTTGTACAGCGCGGTG 99
DB 171 ThrAlaLeusergluileuArgSerGlyLysArgProAspLeuValSerTyValGln 190
QY 100 CTGTGTGCG----- 108
DB 191 ThrLeuCylysglyLeuserGlnProThrhAsnLeuValAlaGlyCylyseuGlnLeu 210
QY 109 -----CAGGCGCAACGAGGAGGATCTTCATGATCGAGCGCGCG 147
DB 211 AsnSerArgAsnPheLeuthrGlnGlnGlyArgAspGlyAla***ArgPheHisGlySer 230
QY 148 AGTCACTTGATATACTTGTGCATGCGCTGCTTTCATATCGCGCGCGCG----- 198
DB 231 Glyly-----ProPheAlaMetHisProtyrProtyrProcyserArgGly 247
QY 199 -----ACTGCTGACGCTGCATACCGAGGCGCTACCTTTCAGCGCGCGGT 243
DB 248 ArgThrValProGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 267
QY 244 CGC---ACCGGTACACACCG---TTGAGAGCTGGCGCGCGCGCGCGCGCGCGCG 297
DB 268 TyrcyAlaAlaAlaTyrglyuthrLeuTyrrAlaAlaAlaGlyGlyGlyAlaSerProAsp 287
QY 298 TATGCGGAGTAGTGTATCAAGCGCGAGGTATGCAATTAATTGTAAGGCGGTGTTAC 357

```

```

Db 288 TyrAsnSerSerGluTyrGluGlyProLeuSerProProLeuCySLeuAsnGlyAsnPhe 307
    ||| |||:|:| |||:|:| |||:|:|
QY 358 GCTGCATACCCCTACGCCCCACCCCTGCGCACTGCTGCTGCTACAGTGACAGTTAC 417
    :|:| :|:| :|:| ||| ||| |||
Db 308 SerLeuLysGlnAspSerSerProAspHisGluLysSerTyrHisTyrSer----- 324
    :|:| :|:| :|:| ||| ||| |||
QY 418 GGACGAGTTATGCTGCCGACCCCTACCACACACACTTGTCTCCAGCCCCCACCCTACGGC 477
    |||:|:|:|:| |||:|:|:|:| |||:|:|:|:|
Db 325 -----MetHisTyrSerGlyCysProGlySerArgHisGly 336
    |||:|:|:|:| |||:|:|:|:| |||:|:|:|:|
QY 478 GTTGGT 483
    ||| |||
Db 337 HisGly 338

```

Search completed: March 12, 2003, 22:21:29
 Job time : 17.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 12, 2003, 21:18:54 ; Search time 15 Seconds
(without alignments)
3424.218 Million cell updates/sec

Title: US-09-809-545A-1_COPY_535_1143

Perfect score: 1164
Sequence: 1 atgaccataaagaagcgctt.....acacacgtttgtccacat 609

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 188354 seqs, 42170167 residues

Total number of hits satisfying chosen parameters: 376708

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+n2p model -DEV=x1h
-Q=/cgn2_1/USFTO.spool/US09809545/runat_11032003_130947_12355/app_query.fasta_1.775
-DB=Published_Applications_AA -QFMT=fastan -SUFITX=n2p.rapb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonsum62
-TRANS=human40.ccd -LIST=45 -DOCALIGN=200 -THR SCORE=pcr -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09809545 @cgn 1.1 9 @runat_11032003_130947_12355
-NCPU=6 -ICPU=3 -NO_XLPRX -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -MAIN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1086	93.3	203	US-09-809-545A-2	Sequence 2, Appli
2	624	53.6	330	US-09-794-591-2	Sequence 2, Appli
3	98.5	8.5	1168	US-09-919-603-2	Sequence 2, Appli
4	96	8.2	760	US-08-754-311B-2	Sequence 2, Appli

5	95.5	8.2	168	US-09-864-761-40976	Sequence 40976, A
6	95.5	8.2	1172	US-09-919-770-4	Sequence 4, Appli
7	95.5	8.2	1172	US-09-822-682-2	Sequence 2, Appli
8	95	8.2	230	US-09-874-062-3	Sequence 3, Appli
9	92	7.9	219	US-09-922-469-2	Sequence 2, Appli
10	92	7.9	219	US-09-922-480-2	Sequence 2, Appli
11	92	7.9	219	US-09-923-236-2	Sequence 2, Appli
12	90.5	7.8	994	US-10-004-551-14	Sequence 14, Appli
13	90.5	7.8	5179	US-10-025-380-1068	Sequence 16, Appli
14	90.5	7.8	5179	US-09-922-217-1068	Sequence 1068, Ap
15	90.5	7.8	5179	US-09-833-263-1068	Sequence 1068, Ap
16	90.5	7.6	171	US-09-764-877-1367	Sequence 1567, Ap
17	88.5	7.5	830	US-09-870-759-134	Sequence 134, App
18	88	7.5	830	US-09-870-759-134	Sequence 140, App
19	88	7.5	196	US-09-989-920-224	Sequence 224, App
20	87	7.5	650	US-09-784-358-10	Sequence 8, Appli
21	87	7.5	724	US-09-784-358-8	Sequence 8, Appli
22	87	7.5	724	US-09-784-358-14	Sequence 14, Appli
23	87	7.5	771	US-09-870-759-134	Sequence 134, App
24	87	7.5	830	US-09-870-759-134	Sequence 140, App
25	87	7.5	830	US-09-784-358-12	Sequence 12, Appli
26	87	7.5	845	US-09-784-358-16	Sequence 16, Appli
27	87	7.5	1617	US-09-784-358-2	Sequence 2, Appli
28	87	7.5	1631	US-09-784-358-2	Sequence 2, Appli
29	87	7.5	1706	US-10-024-450-2	Sequence 659, App
30	86.5	7.4	171	US-09-925-297-659	Sequence 14, Appli
31	86.5	7.4	262	US-10-235-674-14	Sequence 14, Appli
32	86.5	7.4	262	US-09-263-689-14	Sequence 20, Appli
33	86.5	7.4	874	US-09-796-753-20	Sequence 2, Appli
34	86.5	7.4	1274	US-10-020-215-2	Sequence 20, Appli
35	86	7.4	245	US-09-796-858-20	Sequence 20, Appli
36	86	7.4	551	US-09-920-300A-1789	Sequence 1789, Ap
37	86	7.4	551	US-10-033-528-1789	Sequence 1789, Ap
38	86	7.4	1216	US-09-938-330-12	Sequence 12, Appli
39	86	7.4	1222	US-09-938-330-8	Sequence 8, Appli
40	86	7.4	1235	US-09-938-330-16	Sequence 16, Appli
41	86	7.4	1232	US-09-938-330-20	Sequence 20, Appli
42	86	7.4	1535	US-10-189-971-14	Sequence 14, Appli
43	86	7.4	1570	US-10-189-971-12	Sequence 12, Appli
44	85.5	7.3	310	US-10-084-094-12	Sequence 12, Appli
45	85	7.3	231	US-09-864-761-37946	Sequence 37946, A

ALIGNMENTS

RESULT 1
US-09-809-545A-2
; Sequence 2, Application US/09809545A
; Patent No. US20020110804A1
; GENERAL INFORMATION:
; APPLICANT: Stanton, Lawrence W.
; TITLE OF INVENTION: SECRETED FACTORS
; FILE REFERENCE: SCIOS.017A
; CURRENT APPLICATION NUMBER: US/09/809,545A
; CURRENT FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-809-545A-2

Alignment Scores:

Pred. No.:	4.65e-88	Length:	203
Score:	1086.00	Matches:	203
Percent Similarity:	100.00%	Conservative:	0
Best local Similarity:	100.00%	Mismatches:	0
Query Match:	93.30%	Indels:	0
DB:	10	Gaps:	0

US-09-809-545A-1_COPY_535_1143 (1-609) x US-09-809-545A-2 (1-203)

QY 1 ATGACTAATAAAGAGCCGTGAACCCCTACACCAATGCTGGAATTAATCCAGTTGTG 60
DB 1 MetThrAnLysLysAlaValAsnProTyrThrAnGlyTyrLysLeuAsnProValVal 20
QY 61 GCGCGGCTCTACAGCCCGGCACTTCTATGCAAGCAGGTCTGTGTGCGCAAGCCACCG 120
DB 21 GlyAlaValAltySerProAspPheTyrAlaGlyThrValLeuLeuGlnAlaAsnGln 40
QY 121 GAGGATCTTCATGATGACAGTGGCCCGGCACTTGTATATCTTCTGCAATGCTGSC 180
DB 41 GluGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
QY 181 TTTCATATCCGCGCGCCCACTGCTGAGCTGCAATCCAGGGGCTCACCCTGAGGCCGT 240
DB 61 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 80
QY 241 GGTCCACCGCTGTACACACCTTTCAGAGCTGGCGGCCCGCCCAATCCCGGCTAT 300
DB 81 GlyArgThrValTyrAsnThrPheArgAlaAlaAlaProProProIleProAlaTyr 100
QY 301 GCGGAGTAGTGTATCAAGAGCCAGTGTATGGCAATTAATGCTACAGGCTGTTACGCT 360
DB 101 GlyGlyValValTyrGlnGlnProValTyrGlyAsnLysLeuLeuGlnGlyTyrAla 120
QY 361 GCATACCGCTACGCCCGCCACCCCTGCGCACTGCTGCTGCTACAGTACAGTACGGA 420
DB 121 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaTyrSerAspSerTyrGly 140
QY 421 CGAGTTTATGCTGCGGACCCCTACACACACACTTGTCTCCAGCCCGCCACTACGGCTT 480
DB 141 ArgValTyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGlyVal 160
QY 481 GGTGCATGAATGCTTTTGGCCCTTGACCGATGCCAAGTACAGGACCATGCTGATGAT 540
DB 161 GlyAlaMetAsnAlaPheAlaPheLeuThrAspAlaLysThrArgSerHisAlaAspAsp 180
QY 541 GTGGCTCTCTCTTCTTCTTATGACGCTAGTATATACCAAGGGGATACACCGCTTT 600
DB 181 ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrGlnGlyGlyTyrAsnArgPhe 200
QY 601 GCTCCATAT 609
DB 201 AlaProTyr 203

RESULT 2
US-09-794-591-2
; Sequence 2, Application US/09794591
; Patent No. US20010018198A1
; GENERAL INFORMATION:
; APPLICANT: Pulster, Stefan M.
; APPLICANT: Shibata, Hiroki
; TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
; TITLE OF INVENTION: Products Related Thereto and Methods of Using Same
; FILE REFERENCE: CE 3093
; CURRENT APPLICATION NUMBER: US/09/794,591
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/145,391
; PRIOR FILING DATE: 1998-09-01
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-794-591-2

Alignment Scores:
Pred. No.: 2,996-47 Length: 330
Score: 624.00 Matches: 119
Percent Similarity: 89.78% Conservative: 4
Best Local Similarity: 86.86% Mismatches: 12
Query Match: 53.61% Indels: 2

DB: 10 Gaps: 0
US-09-809-545A-1_COPY_535_1143 (1-609) x US-09-794-591-2 (1-330)
QY 1 ATGACTAATAAAGAGCCGTGAACCCCTACACCAATGCTGGAATTAATCCAGTTGTG 60
DB 1 MetThrAnLysLysThrValAsnProTyrThrAnGlyTyrLysLeuAsnProValVal 215
QY 61 GCGCGGCTCTACAGCCCGGCACTTCTATGCAAGCAGGTCTGTGTGCGCAAGCCACCG 120
DB 216 GlyAlaValAltySerProGluPheTyrAlaGlyThrValLeuLeuGlnAlaAsnGln 235
QY 121 GAGGATCTTCATGATGACAGTGGCCCGGCACTTGTATATCTTCTGCAATGCTGSC 180
DB 236 GluGlySerSerMetTyrSerAlaProSerSerLeuValTyrThrSerAlaMetProGly 255
QY 181 TTTCATATCCGCGCGCCCACTGCTGAGCTGCAATCCAGGGGCTCACCCTGAGGCCGT 240
DB 256 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 275
QY 241 GGTCCACCGCTGTACACACCTTTCAGAGCTGGCGGCCCGCCCAATCCCGGCTAT 300
DB 276 GlyArgThrValTyrAsnThrPheArgAlaAlaAlaProProProIleProAlaTyr 295
QY 301 GCGGAGTAGTGTATCAAGAGCCAGTGTATGGCAATTAATGCTTACAGGCTGTTACGCT 360
DB 296 GlyGlyValValTyrProGlyTyrIleLeuTyrCyAspHisLeu--TyrIleLeuCySc 315
QY 361 GCATACCGCTACGCCCGCCACCCCTGCGCACTGCTGCTGCTACAGTACAGTACGGA 407
DB 315 YsIleProLeuArgProAlaTyrProCyHisLysCyArgCysLeuGln 330

RESULT 3
US-09-919-603-2
; Sequence 2, Application US/09919603
; Patent No. US20020137679A1
; GENERAL INFORMATION:
; APPLICANT: Lawler, John W.
; TITLE OF INVENTION: COMP/TSP-1, COMP/TSP-2 and Other TSP
; TITLE OF INVENTION: Chimeric Proteins
; FILE REFERENCE: 1440 1033-007
; CURRENT APPLICATION NUMBER: US/09/919,603
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: PCT/US00/02482
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/118,053
; PRIOR FILING DATE: 1999-02-01
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1168
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-919-603-2

Alignment Scores:
Pred. No.: 0.905 Length: 1168
Score: 98.50 Matches: 50
Percent Similarity: 35.55% Conservative: 25
Best Local Similarity: 23.70% Mismatches: 66
Query Match: 8.46% Indels: 71
DB: 10 Gaps: 13

US-09-809-545A-1_COPY_535_1143 (1-609) x US-09-919-603-2 (1-1168)
QY 12 AAAGCCGCGTAACCCCTACACCAATGCTGGAATTAATCCAGTTGTGCGCGGTCTA 71
DB 380 GluGlyTyrPheProTyrAlaGluThrPheGlnCysSerValThrCysGlySerGlyThr 399
QY 72 GAGCCCGGCT-----TCTATGCGAG 91
DB 400 GlnGlnArgGlyArgSerCysAspValThrSerAsnThrCysLeuGlyProSerIleGln 419

Qy	92	GCACGGGCTCT	-----TGCGCCAGGCCAACGAGAGAT	-----	127
Db	420	ThrArg1AcysSerLeuSerLysCysAspThrArg1IleArg1GlnSerGlyGlyTyrPse	-----	439	
Qy	128	-----	CTTCATGTACAGTGGCCCCAGTTCACTGTAT	160	
Db	440	HisTrpSerProTyrPseSerCysSerValIleThrCysGlyValGlyAsnIle-ThrArg1	-----	459	
Qy	161	ATACTTCTGCAATGCTGGGCTT	-----TCCATATCCGGCCCGCAGCTGTCAGCTGCATACC	217	
Db	459	eArgLeuGlyAsnSerProValProGlnMetGlyLysAsnGly	-----	474	
Qy	218	GAGGGGCTCACCTTGCAGGCCGTGGTGCACACCGTGTAACAACCTTAGAGCTGC	---GG	274	
Db	475	-LysSerLysArgLys	-----ArgGlnThrLysAlaCysGlnG	486	
Qy	275	CGCCCCACCCCAATCCCGGCTATGAGCGAGTAGTGTAACAAGACCGGTATAGCA	334		
Db	486	YAlaProCysProIleAspGlyAlaGlyTrp	-----SerProTyr	498	
Qy	335	ATAAATGTCTACAGAGGTGGTTACCTGCATACCGCTACGCCACCCACTCCACTG	394		
Db	499	-----SerProTyrPseSerAlaCysThrVal	-----Thr	507	
Qy	395	CTGCTGCCTACAGTACACATTACGAGACGATTATGCTGCCAGCCCTACACACACAC	452		
Db	507	rCysAlaGlyLysIleArgGlnArgTrpThrArgValAsnSerProGlnProGlnTyrG	527		
Qy	453	-----ACTTGC	---TCCACGCCCACTACGGCGTGGTGTCATGAATGCTTTGGCG	502	
Db	527	yGlyLysAlaCysValGlyAspValGlnGlnArgGlnMetCysAsnLysArgSerCysPr	547		
Qy	503	CCTTGACCGTAGCCACAGACTAGAGCCATGC	533		
Db	547	oValAspGlyCysLeuSer	---AsnProCys	556	

1 RESULT 4
 2 US-08-754-11B-2
 3 Sequence 2, Application US/08754311B
 4 Patent No. US20020004221A1
 5
 6 GENERAL INFORMATION:
 7 APPLICANT: Bonini, Nancy M.
 8 APPLICANT: Leiserson, William M.
 9 APPLICANT: Benzer, Seymour
 10 TITLE OF INVENTION: PROGRAMMED CELL DEATH ANTAGONIST
 11 TITLE OF INVENTION: PROTEINS
 12 NUMBER OF SEQUENCES: 8
 13 CORRESPONDENCE ADDRESS:
 14 ADDRESSEE: Flehr, Hobbach, Test, Albrighton & Hebert
 15 STREET: 4 Embarcadero Center, Suite 3400
 16 CITY: San Francisco
 17 STATE: California
 18 COUNTRY: United States
 19 ZIP: 94111-4187
 20
 21 COMPUTER READABLE FORM:
 22 MEDIUM TYPE: Floppy disk
 23 COMPUTER: IBM PC compatible
 24 OPERATING SYSTEM: PC-DOS/MS-DOS
 25 SOFTWARE: PatentIn Release #1.0, Version #1.25
 26
 27 CURRENT APPLICATION DATA:
 28 APPLICATION NUMBER: US/08/754,311B
 29 FILING DATE: 21-NOV-1996
 30 CLASSIFICATION: 435
 31 PRIOR APPLICATION DATA:
 32 APPLICATION NUMBER: US 08/195,152
 33 FILING DATE: 14-FEB-1994
 34
 35 ATTORNEY/AGENT INFORMATION:
 36 NAME: Treacartin, Richard F.
 37 REGISTRATION NUMBER: 31,801
 38 REFERENCE/DOCKET NUMBER: A-59551/RFT/RWS
 39 TELECOMMUNICATION INFORMATION:
 40 TELEPHONE: (415) 781-1889
 41 TELEFAX: (415) 398-3249

```

; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 760 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-754-311B-2

Alignment Scores:
Pred. No.: 1.38
Score: 96.00
Percent Similarity: 37.29%
Best Local Similarity: 23.73%
Query Match: 8.25%
DB: 8

length: 760
Matches: 42
Conservative: 24
Mismatches: 61
Indels: 50
Gaps: 6

```

US-09-809-545A-1_COPY_535_1143 (1-609) x US-08-754-311B-2 (1-760)

OY 124 GGATCTTCATGATAC-----AGTGGCCCCAGTTCACTT 156
Db 223 GlySerAsnLeuTyrGlyCysSerSerAlaSerAsnProLeuAspGlyAlaValAla 242	
157 GATATACCTTCGCAGATGCTGGCTTTCATATCCGGCGCCACTGGCTGCAGCTGATAC 216	
Db 243 ValAsnSerSerAlaVal-----AlaAlaAlaAlaAlaValTyr 256	
217 CGAGGGGCTCATCTTCGAGCGCGTGTGCACCCGTGCATACACCTTCAGAGCTGCGCG 276	
Db 257 AspGlyLysHis-----AspTyrTyrTyrTyrAsnSerMetGlnGlnTyrThr 272	
277 CCCCACACCCCAATCCCGGCTATGCGCGAGTATGTATATACAGCAGTATAGCAAT 336	
Db 273 PropProProPheTyrSerGlyTyrGlyThrProTyrAlaAlaAlaThrAlaAlaArgGln 292	
337 AATATGCTACAGAGGTGGTACGCTGCATACCGCTACGCCCGACCCCTGCACATGCT 396	
Db 293 AlaLysMetGlnProGlyAlaAlaAlaAlaAlaAlaTyrLeuThrProSerTyrAla 312	
397 GCTGCC----- 402	
Db 313 AlaSerGlyAsnAsnAsnSerGlnLeuTyrSerSerProTyrAlaGlyTyrAsnAsnPhe 332	
403 -----TACAGTACAGATTACGAGCAGATTATGCTGCCGAC 438	
333 GlyGlnGlnAspTyrGlyTyrTyrTyrAsnGlnGlnTyrGlyAsnTyrTyrSerProAla 352	
439 CCTTACCCACACACACTTCTCAGCCCC-----ACCTACGGGGTGTGT--- 483	
Db 353 AsnTyrSerProTyrAlaValSerSerProSerSerSerAlaSerHisGlyHisGlyPhe 372	
484 GCCATGATATCTTTGGCCCTTGACCGATGCCAGACATGAGCCATGCT 534	
Db 373 HisValAlaAlaSerSerAsnLeuSerGlnUserProThrAspThrHisSer 389

RESULT 5
US-09-864-761-40976
; Sequence 40976, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26

[illegible]

```

US-09-919-770-4
Result 6
Sequence 4, Application US/09919770
Patent No. US20020046577A1
GENERAL INFORMATION:
APPLICANT: Bornstein, Paul
APPLICANT: Kyriakides, Themis
APPLICANT: Ratner, Buddy
APPLICANT: Giachelli, Cecilia
APPLICANT: Martinson, Laura
APPLICANT: Scatena, Maria
TITLE OF INVENTION: Methods and Devices to Modulate the Wound Response
FILE REFERENCE: US09919770
CURRENT APPLICATION NUMBER: US/09/919,770
CURRENT FILING DATE: 2001-07-31
PRIORITY FILING DATE: 2000-08-01
PRIORITY FILING DATE: 2000-08-01
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 1172
TYPE: RTT
ORGANISM: Homo Sapien
US-09-919-770-4

```

US-09-809-545A-1_COPY_535_1143 (1-609) x US-09-919-770-4 (1-1172)

OY	12	AAAGCCGTGAACCCCTCAACCAATGCGTGGAAATTAAATTCAGATTGCGCGGCTA	71
Db	382	GtnglyTrpSerProtpalaGlutrpThcInCysSerValtnrCysGlySerglyThr	401
OY	72	CAGCCCCGACT-----TTATGCAG	91
Db	402	GlnGlnArgglyArgSerCysAspValtnrSeranthrCysLeughlyProserIleGln	421
OY	92	GCAGGCTCTGT-----TGTCGACGCCAACAGGAGGAT-----	121
Db	422	ThrtAlaCysSerLeuSerLysCysAspThrArglleArgGlnAspGlyglyTyrSer	441
OY	128	-----CTTCACGTACAGGTGCCGCCAGTCACTTGTAT	166
Db	442	HistTrpSerProtpSerSerCysSerValtnrCysglValGlyAsnIle-ThrtArgil	461
OY	161	ATACTTCTGCATGCTGGCT--TCCATATCCGCGCCGACACTGCTGCACGTGCATACC	217

```

Db 461 eArygleuCyAsnSerProValProGlnMeGlyGlyLysAsnCyS----- 476
Qy 218 GAGGGGCTCACTTCGAGCGCGTGTGCACCGGTACACACCTTCAGAGCTC---GG 274
Db 477 -LysGlySerGly-----ArgGlnThrLysAlaCySgIngl 488
Qy 275 CGCCCCACCCCAATCCCGCGCTATGCGAGTAGTGATCAAGACCGATGATGCA 334
Db 488 yAlaProCySProlleAspGlyArgTrp-----SerProTrp-- 500
Qy 335 ATAAATGCTACAGGGGTTCACCTACATCCGCTACGCCACCCACCTTGCCACTG 394
Db 501 -----SerProTrpSerAlaCySThrVal-----Th 509
Qy 395 CTGCTGCTACAGTACAGTTCACGAGATTATGCTTACCGACCCCTACACACAC-- 452
Db 509 rCySaAlaGlyGlyLleArgGlnArgThrArgValCySaAsnSerProGlnProGlnTrpG 529
Qy 453 -----ACTTGC---TCCAGCCCCCACCCTACGCGCTGTGCTGATGATGCTTTGCGC 502
Db 529 yGlyLysAlaCySaValGlyAspValGlnGlnArgGlnMeCySaSlnLysArgSerCySP 549
Qy 503 CCTTGACCCGATGCCCAAGACTAGAGCCATGC 533
Db 549 oValAspGlyCySLeuSer---AsnProCyS 558

```

RESULT 7

```

US-09-822-682-2
; Sequence 2, Application US/09822682
; Patent No. US20020119921A1
; GENERAL INFORMATION:
; APPLICANT: Detmar, Michael J.
; TITLE OF INVENTION: THROMBOSPONDIN-2 AND USES THEREOF
; FILE REFERENCE: 10287-051002
; CURRENT APPLICATION NUMBER: US/09/822,682
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/536,087
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/127,221
; PRIOR FILING DATE: 1999-03-31
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1172
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-822-682-2

```

Alignment Scores:

```

Pred. No.: 1 67 Length: 1172
Score: 95.50 Matches: 50
Percent Similarity: 35.55% Conservative: 25
Best Local Similarity: 23.70% Mismatches: 66
Query Match: 8.20% Indels: 71
DB: 10 Gaps: 13

```

```

US-09-809-545A-1_COPY_535_1143 (1-609) x US-09-822-682-2 (1-1172)
Qy 12 AAAGCCGTGAACCCCTACCAATGCTGGAAATTAATCAGTTGTGGCGCGGTCTA 71
Db 382 GlnGlyTrpSerProTrpAlaGlnTrpThrGlnCySserValThrCySgInglThr 401
Qy 72 CAGCCCGGACT-----TCTATGAG 91
Db 402 GlnGlnArgGlyArgSerCySaSPValThSerAnThrCySLeuGlyProSerIleGln 421
Qy 92 GCACGGGCGGT-----TGTCACAGGCCAACGAGGAGGAT----- 127
Db 422 ThrArgAlaCySserLeuSerLysCySaSPThrArgIleArgLysAspGlyTrpSer 441
Qy 128 -----CTTCATGATACAGTGGCCGACGTTCACTTGAT 160

```

```

Db 442 HisTrpSerProTrpSerSerCySserValThrCySglValGlyAsnIle-ThrArgI 461
Qy 161 ATACTTTCGATGACCTGCGCT---TCCATATCCGGCGGCGGCTGCTGACATACC 217
Db 461 eArygleuCyAsnSerProValProGlnMeGlyGlyLysAsnCyS----- 476
Qy 218 GAGGGGCTCACTTCGAGCGCGTGTGCACCGGTACACACCTTCAGAGCTC---GG 274
Db 477 -LysGlySerGly-----ArgGlnThrLysAlaCySgIngl 488
Qy 275 CGCCCCACCCCAATCCCGCGCTATGCGAGTAGTGATCAAGACCGATGATGCA 334
Db 488 yAlaProCySProlleAspGlyArgTrp-----SerProTrp-- 500
Qy 335 ATAAATGCTACAGGGGTTCACCTACATCCGCTACGCCACCCACCTTGCCACTG 394
Db 501 -----SerProTrpSerAlaCySThrVal-----Th 509
Qy 395 CTGCTGCTACAGTACAGTTCACGAGATTATGCTTACCGACCCCTACACACAC-- 452
Db 509 rCySaAlaGlyGlyLleArgGlnArgThrArgValCySaAsnSerProGlnProGlnTrpG 529
Qy 453 -----ACTTGC---TCCAGCCCCCACCCTACGCGCTGTGCTGATGATGCTTTGCGC 502
Db 529 yGlyLysAlaCySaValGlyAspValGlnGlnArgGlnMeCySaSlnLysArgSerCySP 549
Qy 503 CCTTGACCCGATGCCCAAGACTAGAGCCATGC 533
Db 549 oValAspGlyCySLeuSer---AsnProCyS 558

```

RESULT 8

```

US-09-874-062-3
; Sequence 3, Application US/09874062
; Patent No. US20020081607A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Four Disulfide Core Domain-Containing (FDCD) Polynucleotides,
; FILE REFERENCE: PT033P1
; CURRENT APPLICATION NUMBER: US/09/874,062
; CURRENT FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: PCT/US00/32462
; PRIOR FILING DATE: 2000-11-29
; PRIOR APPLICATION NUMBER: 60/168,229
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-874-062-3

```

Alignment Scores:

```

Pred. No.: 1 32 Length: 230
Score: 95.00 Matches: 56
Percent Similarity: 32.27% Conservative: 15
Best Local Similarity: 25.45% Mismatches: 65
Query Match: 8.16% Indels: 86
DB: 10 Gaps: 13

```

```

US-09-809-545A-1_COPY_535_1143 (1-609) x US-09-874-062-3 (1-230)
Qy 54 AGTTTGCGCGCGGCTTACA-----GCCCGACTTTCATGAGGACGG 97
Db 14 SerCySglYglnGlyLeuGlnGlnArgSerCySaArgProSerCySgInThrThr 33
Qy 98 TGCTGT-----TGTCACAGGCCAACGAGGAGATTCATGATACAGGCGCCCA 148
Db 34 CySglArgGThrThrCySArgProSer-----Cys----- 44
Qy 149 GTTCACTTGTATATCTTCGATGCGCTGCTTCATATCCGCGCGGACGCTGCG-- 206

```


Db 87 -----ValTyr-----HisIleArgGlyPheProLeuAlaThrGlnLeuAsnVa 101
Qy 238 -----CGTGTCCGACCGGTGAC-----AACACCTTCAGAGCTGC 272
Db 101 lProProLeuProProArgGlyPheProPheValProProSerArgPhePheSerAlaAl 121
Qy 273 GGGGCCCCCA-----CCCCCAATCCGGCCTTAGCGGAGTACTGATCAAGCCAGT 326
Db 121 aaIaAlaProAlaAlaProProIleAlaAla-----GluProAl 134
Qy 327 GTATGCAATAAATTTGCTACAGGGGTGTACGCTGCATACCGCTACGCCCAACCC 386
Db 134 a-----AlaAlaAlaProLeuThr 140
Qy 387 TGCCACTGCTGCTCCTACAGTACGATGACGAGATTATGCTGCGACCCCTACCA 446
Db 140 rAlaThrProValAlaAlaGluProAlaAlaGlyAlaProValAlaAlaGluPro----- 158
Qy 447 CCACACACTTGCTCCAGCCCCCACCCTACGGCGTTGTGCC 486
Db 159 -----AlaAlaGluAlaPro-----ValGlyAla 166
RESULT 11
US-09-923-236-2
; Sequence 2, Application US/09923236
; Patent No. US20020090677A1
; GENERAL INFORMATION:
; APPLICANT: Shepherd, Paul O.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/923,236
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-923-236-2
Alignment Scores:
Pred. No.: 2.4 Length: 219
Score: 92.00 Matches: 53
Percent Similarity: 33.91% Conservative: 6
Best Local Similarity: 30.46% Mismatches: 56
Query Match: 7.90% Indels: 59
Gaps: 10
US-09-809-545A-1_COPY_535_1143 (1-609) x US-09-923-236-2 (1-219)
Qy 24 CCCCTACCAATGCTGGAATTAATCCAGTTGTGGCGCGGTCTACAGCCCGACTT 83
Db 33 ProLeuHisProSerLeuAenIleProTyrGlyIle---ArgaenLeuProProLeu 51
Qy 84 CTATGACGGCAGCGGTGCTTTGGCCAGGCCAACCCAGAGGAGGATCTTCATGACAGTG 143
Db 52 TyrTyrArgProValaenThrValaProSerTyrProGlyAenThrTyr---ThraepThrG 71
Qy 144 CCCAGTTCATTTATATATCTTCGCAATGCTGGCTTCATATCCGCGCGGCACAGC 203
Db 71 yLeuProSerTyrProTrrIleLeuThrSerProGlyPheProTyr----- 86
Qy 204 TGCAGTGCATACCGAGGGGCTCACTTCAGGC----- 237
Db 87 -----ValTyr-----HisIleArgGlyPheProLeuAlaThrGlnLeuAsnVa 101
Qy 238 -----CGTGTCCGACCGGTGAC-----AACACCTTCAGAGCTGC 272
Db 101 lProProLeuProProArgGlyPheProPheValProProSerArgPhePheSerAlaAl 121

Qy 273 GGGGCCCCCA-----CCCCCAATCCGGCCTTAGCGGAGTACTGATCAAGCCAGT 326
Db 121 aaIaAlaProAlaAlaProProIleAlaAla-----GluProAl 134
Qy 327 GTATGCAATAAATTTGCTACAGGGGTGTACGCTGCATACCGCTACGCCCAACCC 386
Db 134 a-----AlaAlaAlaProLeuThr 140
Qy 387 TGCCACTGCTGCTCCTACAGTACGATGACGAGATTATGCTGCGACCCCTACCA 446
Db 140 rAlaThrProValAlaAlaGluProAlaAlaGlyAlaProValAlaAlaGluPro----- 158
Qy 447 CCACACACTTGCTCCAGCCCCCACCCTACGGCGTTGTGCC 486
Db 159 -----AlaAlaGluAlaPro-----ValGlyAla 166
RESULT 12
US-10-004-551-14
; Sequence 14, Application US/10004551
; Publication No. US20030004310A1
; GENERAL INFORMATION:
; APPLICANT: SHIMKETS, RICHARD A
; APPLICANT: FERNANDES, EUMA
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: 15966-559
; CURRENT APPLICATION NUMBER: US/10/004,551
; CURRENT FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 09/635,949
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 993
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION: n 1755 can be A, G, C, or T
US-10-004-551-14
Alignment Scores:
Pred. No.: 4.44 Length: 993
Score: 90.50 Matches: 57
Percent Similarity: 38.27% Conservative: 36
Best Local Similarity: 23.46% Mismatches: 79
Query Match: 7.77% Indels: 72
Gaps: 12
US-09-809-545A-1_COPY_535_1143 (1-609) x US-10-004-551-14 (1-993)
Qy 34 AATGGCTGGAATTAATCA-----GTTGGGCGCGGTCTACAGCCCG 78
Db 718 AsnGlyTrrPlys---SerProSerGlnProGluLeuValHisGlyThrValThrTyr 736
Qy 79 GACTTCTATGACAGC-----ACGGTGTGTTGCCAGC----- 113
Db 737 GlnCysTyrProGlyTyrGlnValaIaGlySerSerValaIleuMetCysGlnTrrPaspLeu 756
Qy 114 -----CAACGAGAGGG 125
Db 757 ThrTrrSerGlnAspLeuProSerCysGlnArgValThrSerCysHisAspProGlyAsp 776
Qy 126 ATCTTCAT-----GTACAGTGGCCCGACTTC-----ACTTGATA 161
Db 777 ValGlnHisSerArgArgGluIleSerSerProLysPheProValaGlyAlaThrValaGln 796
Qy 162 TACTTCTGCAATGCTGCTTCATATCCGCGCCACACTGCTGCAGCTGCATACCGAGG 221
Db 797 TyrIleCysaspGlnGlyPheValaIleuThrGlySerSerIleLeuThrCysHisAspArg 816
Qy 222 GGCCTACCTTCAGAGCGCGTGTGGCACCGT-----GTACAAACACTTCAG 266
Db 817 GlnAlaGlySerProLysTrrPrrSerAspArgAlaProLysCysLeuLeuGlnLeuLys 836

```
QY 267 AGTCG-----GGGCCCCCACCACCCCAATCCGGCCTATGG 302
    |||
Db 837 ProCysHISglYleuSerAlaProGluAnGlyAlaArgSerProGluGlnLeuHis 856
QY 303 CGGAGTAgTATCAAGAGCCAGTATAGGCAATAAATTGCTACAGGGTGG----- 353
    |||
Db 857 ProAlaGly-----AlaThrIleHisPheSerCysAlaProGlyTyrValLeuLys 873
QY 354 -----TTACGGCTGATACCGGCTACCGCCAC-----CCC 386
    |||
Db 874 GlyAlaIaSerIleLysCysValProGlyHisProSerHisIleTyrSerAspProPro 893
QY 387 TGCCACTGCTGCTGCTACAGTACAGTACGAGCAAGTTATAGTCCGACCCCTACCA 446
    |||
Db 894 -IleCysArgAlaIaIaSerLeuAspGlyPhe-----TyrAsnSerArgSerLeuAs 910
QY 447 CCACACACTTGTCTCAAGCCCACTACGAGCCTGTGTCATGATGCTTTGGCCCTT 506
    |||
Db 910 pValaIaLysAlaProAlaIaIaSerSerThrLeuAspAlaAlaHisIleAlaIaIa 930
QY 507 GACCGATGCAAGACTAGAGAGCAGTATGATGATGGGTCTCGTTCTTTCTTCAATGCA 566
    |||
Db 930 ePhe-LeuProLeuValAlaMetValLeuValGlyGlyValTyrThrIlePheSerA 950
QY 567 GGCTA 571
    |||
Db 950 rGleu 951

RESULT 13
US-10-004-551-16
: Sequence 16, Application US/10004551
: Publication No. US20030004310A1
: GENERAL INFORMATION:
: APPLICANT: SHIMKETS, RICHARD A
: APPLICANT: FERNANDES, ELMA
: TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
: FILE REFERENCE: 15966-559
: CURRENT APPLICATION NUMBER: US/10/004,551
: PRIOR FILING DATE: 2001-12-05
: PRIOR APPLICATION NUMBER: 09/635,949
: NUMBER OF SEQ ID NOS: 110
: SOFTWARE: Patent In Ver. 2.1
: SEQ ID NO 16
: LENGTH: 994
: TYPE: PRT
: ORGANISM: Homo sapiens
: OTHER INFORMATION: n 1755 can be A, G, C, or T.
US-10-004-551-16

Alignment Scores:
Pred. No.: 4 45 Length: 994
Score: 90.50 Matches: 57
Percent Similarity: 38.27% Conservative: 36
Best Local Similarity: 23.46% Mismatches: 79
Query Match: 7.77% Indels: 72
DB: 9 Gaps: 12

US-09-809-545a-1_copy_535_1143 (1-609) x US-10-004-551-16 (1-994)
QY 34 AATGGCTGGAATTAATCA-----GTTGTGGCGCGGTCTTACAGCCCC 78
    |||
Db 718 AanglyItrPlye---serProserGlnProGluLeuValHisIleGlyThrValIaThrTyr 736
QY 79 GACTTCTATGCAAGC-----ACGGTGTCTGTGGCCAGGC----- 113
    |||
Db 737 GInCyetyrProGlyTyrGlnValValGlySerSerValLeuMetCysGlnItrPAspLeu 756
QY 114 -----CAACGAGGAGG 125
    |||
Db 757 ThrItrPserGluAspLeuProSerCysGlnArgValIleSerCysHisAspProGlyAsp 776
QY 126 ATCTTCAT-----GTACAGTGGCCCCAGTTC-----ACTGTATA 161
```

```
Db 777 ValGlnHisSerArgValGluLeuIleSerSerProLysPheProValGlyAlaThrValGln 796
    |||
QY 162 TACTTTCGATAGCCTGCTGCTTTCCATACCGCGCCGACCTGCTGACCTGATACGAGG 221
    |||
Db 797 TyrIleCysAspGlnGlyPheValLeuThrGlySerSerIleLeuThrCysHisAspArg 816
QY 222 GGTACACTTTCGAGCGCGTGTGCTGACACCT-----GTACAACACTTCAG 266
    |||
Db 817 GlnIaGlySerProLysItrPserAspArgAlaProLysCysLeuLeuGlnGlnLeuLys 836
QY 267 AGTCG-----GGGCCCCCACCACCCCAATCCGGCCTATGG 302
    |||
Db 837 ProCysHISglYleuSerAlaProGluAnGlyAlaArgSerProGluLysGlnLeuHis 856
QY 303 CGGAGTAgTATCAAGAGCCAGTATAGGCAATAAATTGCTACAGGGTGG----- 353
    |||
Db 857 ProAlaGly-----AlaThrIleHisPheSerCysAlaProGlyTyrValLeuLys 873
QY 354 -----TTACGGCTGATACCGGCTACCGCCAC-----CCC 386
    |||
Db 874 GlyAlaIaSerIleLysCysValProGlyHisProSerHisIleTyrSerAspProPro 893
QY 387 TGCCACTGCTGCTGCTACAGTACAGTACGAGCAAGTTATAGTCCGACCCCTACCA 446
    |||
Db 894 -IleCysArgAlaIaIaSerLeuAspGlyPhe-----TyrAsnSerArgSerLeuAs 910
QY 447 CCACACACTTGTCTCAAGCCCACTACGAGCCTGTGTCATGATGCTTTGGCCCTT 506
    |||
Db 910 pValaIaLysAlaProAlaIaIaSerSerThrLeuAspAlaAlaHisIleAlaIaIa 930
QY 507 GACCGATGCAAGACTAGAGAGCAGTATGATGATGGGTCTCGTTCTTTCTTCAATGCA 566
    |||
Db 930 ePhe-LeuProLeuValAlaMetValLeuValGlyGlyValTyrThrIlePheSerA 950
QY 567 GGCTA 571
    |||
Db 950 rGleu 951

RESULT 14
US-10-025-380-1068
: Sequence 1068, Application US/10025380
: Publication No. US20020182191A1
: GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: APPLICANT: Lodes, Michael J.
: APPLICANT: Secrist, Heather
: APPLICANT: Benson, Darin R.
: APPLICANT: Meagher, Madeleine Joy
: APPLICANT: Stolk, John A.
: APPLICANT: Wang, Tongtong
: APPLICANT: Jiang, Yugu
: APPLICANT: Smith, Carole L.
: APPLICANT: King, Gordon E.
: APPLICANT: Wang, Aijun
: APPLICANT: Clapper, Jonathan D.
: APPLICANT: Skeiky, Yasir A. W.
: APPLICANT: Fanger, Gary R.
: APPLICANT: Vedrick Thomas S.
: APPLICANT: Carter, Darick
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
: FILE REFERENCE: 210121.471C14
: CURRENT APPLICATION NUMBER: US/10/025,380
: PRIOR FILING DATE: 2001-12-19
: NUMBER OF SEQ ID NOS: 1129
: SOFTWARE: fastseq For Windows Version 4.0
: SEQ ID NO 1068
: LENGTH: 5179
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-025-380-1068
```

Alignment Scores:
 Pred. No.: 6.24 Length: 5179
 Score: 90.50 Matches: 38
 Percent Similarity: 33.83% Conservative: 7
 Best Local Similarity: 28.57% Mismatches: 43
 Query Match: 7.77% Indels: 45
 DB: 9 Gaps: 6

US-09-809-545a-1_copy_535_1143 (1-609) x US-10-025-380-1068 (1-5179)

QY 141 TGGCCCGAGTTCATGTATATCTTGCAGATGCT-----GGCTTCCATATCCGGC 194
 DB 1390 TTPrometalsplyscysile-thrthrProserProthThrThrProserPro-P 1409
 QY 195 CGCCACTGTGCAGCTCATACGAGGGGCTCACCTTCGAGCCGGTGCACCGTGA 254
 DB 1409 roPro-----ThrThrThrThrThrleuPreProthrt 1420
 QY 255 CAACACCTTCA-----GAGCTGGCGGCGCCCGCCCAATCCCGGCTATGG 302
 DB 1420 hrThrProserProProthThrThrThrThrProProProthThrThrProserP 1440
 QY 303 CGAGTGTGTATCAAGAGCCAGTGTATGGCAATTAATGCTACAGGTGTACG---- 358
 DB 1440 roPro-----1leThrThrThrt 1446
 QY 359 -----CTGCATACCGCTACGCCCGCAGCCCGCTGCACCTGCTGCTG----- 400
 DB 1446 hrThrProleuProthThrThrProserProProIleSerThrThrThrProProP 1466
 QY 401 -----CTTACAGTGAAGTTACGAGAGGAGTTATGCTGCGAGCC 440
 DB 1466 roThrThrThrProserProProthThrThrThrProserProProthThrThrProserP 1486
 QY 441 CTACACACACACACTTGCTCCAGCCCGCCACTAGC 475
 DB 1486 roProThrThrThrThrThrThrProProProth 1497

RESULT 15

US-09-922-217-1068
 ; Sequence 1068, Application US/09922217
 ; Patent No. US2002007641A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Lodes, Michael J.
 ; APPLICANT: Secrist, Heather
 ; APPLICANT: Benson, Darin R.
 ; APPLICANT: Meagner, Madeline Joy
 ; APPLICANT: Stolk, John A.
 ; APPLICANT: Wang, Tonglong
 ; APPLICANT: Jiang, Yugu
 ; APPLICANT: Smith, Carole Lynn
 ; APPLICANT: King, Gordon E.
 ; APPLICANT: Wang, Aljun
 ; APPLICANT: Clapper, Jonathan D.
 ; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
 ; FILE REFERENCE: 210121.471C13
 ; CURRENT APPLICATION NUMBER: US/09/922.217
 ; CURRENT FILING DATE: 2001-08-03
 ; NUMBER OF SEQ ID NOS: 1124
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1068
 ; LENGTH: 5179
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-922-217-1068

Alignment Scores:
 Pred. No.: 6.24 Length: 5179
 Score: 90.50 Matches: 38
 Percent Similarity: 33.83% Conservative: 7
 Best Local Similarity: 28.57% Mismatches: 43

Query Match: 7.77% Indels: 45
 DB: 10 Gaps: 6

US-09-809-545a-1_copy_535_1143 (1-609) x US-09-922-217-1068 (1-5179)

QY 141 TGGCCCGAGTTCATGTATATCTTGCAGATGCT-----GGCTTCCATATCCGGC 194
 DB 1390 TTPrometalsplyscysile-thrthrProserProProthThrThrProserPro-P 1409
 QY 195 CGCCACTGTGCAGCTCATACGAGGGGCTCACCTTCGAGCCGGTGCACCGTGA 254
 DB 1409 roPro-----ThrThrThrThrThrleuPreProthrt 1420
 QY 255 CAACACCTTCA-----GAGCTGGCGGCGCCCGCCCAATCCCGGCTATGG 302
 DB 1420 hrThrProserProProthThrThrThrThrProProProthThrThrProserP 1440
 QY 303 CGAGTGTGTATCAAGAGCCAGTGTATGGCAATTAATGCTACAGGTGTACG---- 358
 DB 1440 roPro-----1leThrThrThrt 1446
 QY 359 -----CTGCATACCGCTACGCCCGCAGCCCGCTGCACCTGCTGCTG----- 400
 DB 1446 hrThrProleuProthThrThrProserProProIleSerThrThrThrProProP 1466
 QY 401 -----CTTACAGTGAAGTTACGAGAGGAGTTATGCTGCGAGCC 440
 DB 1466 roThrThrThrProserProProthThrThrThrProserProProthThrThrProserP 1486
 QY 441 CTACACACACACACTTGCTCCAGCCCGCCACTAGC 475
 DB 1486 roProThrThrThrThrThrThrProProProth 1497

Search completed: March 12, 2003, 22:22:08
 Job time : 21 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: March 15, 2003, 22:52:53 ; Search time 1912 Seconds
(without alignments)
3089,890 Million cell updates/sec

Title: US-09-809-545A-2
Perfect score: 1086
Sequence: 1 MTRKKAHPYTNMGKLNPPV.....VLSSLAQSIYGGYRFPAPY 203

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 segs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+D2n.model -DEV=xlp
-Q=/cgn21/USPTO.spool/US09809545/rnuc_07032003.153853.11454/apd_query.fasta_1.391
-DB=GenEmbl -OPMT=faetap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=00 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US09809545 @CCN 1 1 3745 @runat: 07032003.153853.11454 -NCPU=6 -ICPU=3
-NO ALPXY -NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_hca:*
2: gb_hcg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_or:*
21: em_om:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rtd:*
36: em_hcg_man:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcg_hum:*
40: em_hcg_mus:*
41: em_hcg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1086	100.0	1340	6 AX268800	AX268800 Sequence
2	1063	97.9	2000	9 AF229057	AF229057 Homo sapi
3	1061	97.7	1586	10 AF191501	AF191501 Mus muscu
4	1024.5	94.3	1363	10 AF107204	AF107204 Mus muscu
5	1021.5	94.1	3348	9 AF109106	AF109106 Homo sapi
6	1020.5	94.0	2002	10 AB041596	AB041596 Mus muscu
7	1009.5	93.0	1513	9 AK001027	AK001027 Homo sapi
8	1009.5	93.0	2279	9 AF107203	AF107203 Homo sapi
9	968	89.1	1475	9 AB060859	AB060859 Macaca fa
10	955	87.9	2372	6 AR134676	AR134676 Sequence
11	912	84.0	1547	9 AF094849	AF094849 Homo sapi
12	526	48.4	1393	9 AY072786	AY072786 Homo sapi
13	526	48.4	1555	9 BC025281	BC025281 Homo sapi
14	526	48.4	1925	9 AF229058	AF229058 Homo sapi
15	525	48.3	1134	10 AF387322	AF387322 Mus muscu
16	525	48.3	1695	10 AF229055	AF229055 Mus muscu
17	522	48.1	1538	9 BC013115	BC013115 Homo sapi
18	519	47.8	1558	9 AK055213	AK055213 Homo sapi
19	470.5	43.3	1721	10 BC027263	BC027263 Mus muscu
20	467.5	43.0	1876	9 HSP38A20	HSP38A20
21	456	42.0	959	10 AF229056	AF229056 Mus muscu
22	347	32.0	1623	9 HSM803066	AL113700 Homo sapi
23	294	27.1	428	6 AX336745	AX336745 Sequence
24	266.5	24.5	164538	9 AC005774	AC005774 Homo sapi
25	262	24.1	159565	10 AL581075	AL581075 Mouse DNA
26	257	23.7	251	9 HSFOX09	AF109115 Homo sapi
27	253	23.3	209461	2 AC079359	AC079359 Homo sapi
28	242.5	22.3	208632	2 AC073624	AC073624 Homo sapi
29	231.5	21.3	2397	9 AK091695	AK091695 Homo sapi
30	202	18.6	2971	9 HSFOX14	AF109120 Homo sapi
31	202	18.6	115860	2 AC120661	AC120661 Rattus no
32	193	17.8	596	9 HSA333688	AF333688 Homo sapi
33	193	17.8	617	9 HSA335811	AF335811 Homo sapi
34	193	17.8	653	9 HSA335616	AF335616 Homo sapi
35	193	17.8	675	9 HSA334013	AF334013 Homo sapi
36	167	15.4	203649	10 AL603843	AL603843 Mouse DNA
37	164	15.1	108315	9 HS41P2	AL049748 Human DNA
38	164	15.1	156820	2 AC123391	AC123391 Rattus no
39	145	13.4	73929	2 AC103207	AC103207 Rattus no
40	140	12.9	234	9 HSFOX13	AF109119 Homo sapi
41	133	12.2	374	9 HSFOX08	AF109114 Homo sapi
42	125	11.5	156549	9 AC079072	AC079072 Homo sapi
43	123.5	11.4	7503	3 CCA251919	AU251919 Cerialtis
44	118.5	10.9	9704	10 MMU25652	U25652 Mus musculu
45	118.5	10.9	118496	2 AC123318	AC123318 Rattus no

RESULT 1

ALIGNMENTS

AX268800 AX268800 1340 bp DNA linear PAT 29-OCT-2001
LOCUS AX268800
DEFINITION Sequence 1 from Patent WO0174901.
ACCESSION AX268800
VERSION AX268800.1 GI:16541860
KEYWORDS
SOURCE
ORGANISM Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
AUTHORS Stanton, L.W. and White, R.T.
TITLE Secreted factors
JOURNAL Patent: WO 0174901-A 1 11-OCT-2001;
Scienc Inc. (US)
FEATURES
source location/Qualifiers
1..1340
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
BASE COUNT 366 a 394 c 317 g 263 t
ORIGIN
Alignment Scores:
Pred. No.: 1,6e-71 Length: 1340
Score: 1086.00 Matches: 203
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
US-09-809-545A-2 (1-203) x AX268800 (1-1340)
Oy 1 MetThrAsnLysLysAlaValAsnProTyrThrAsnGlyTrpLysLeuAsnProValVal 20
Db 535 ATGACTAATAAAGAGCCGCTGACCCCTACACCAATGCTGGAATTAATCCAGTTG 594
Oy 21 GAlaValAlaTyrSerProAspPheTyrAlaGlyThrValLeuLeuCyseGlnAlaAsnGln 40
Db 595 GCGCGGCTTACAGCCCGCCGCTTCTATGACGACGCGTGTGTGCGCAGCCACAG 654
Oy 41 GAlGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
Db 655 GAGGGATCTTCCATGATGACGTGGCCCGCAGTTCCTGTATATCTCTGCAATGCTGGC 714
Oy 61 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 80
Db 715 TTTCATATCCGGCCCGCCACTGCTGACGCTGACACGAGGGCTCACCTTCAGGCGCT 774
Oy 81 GAlArgThrValTyrAsnThrPheArgAlaAlaAlaProProProProIleProAlaTyr 100
Db 775 GGTGCGACCGTGACAAACCTTCAGAGCTGGCGGCCCGCCACCCCAATCCCGGCTAT 834
Oy 101 GAlGlyValAlaTyrGlnGluProValTyrGlyAsnLysLeuGlnGlyGlyTyrAla 120
Db 835 GCGCGAGTGTATGATCAAGAGCCAGTGTATGGCAATTAATTCCTACAGGGGTGCTACGCT 894
Oy 121 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaTyrSerAspSerTyrGly 140
Db 895 GCATACCGCTACGCCCGACGCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 954
Oy 141 ArgValTyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGlyVal 160
Db 955 CGAGTTATGCTGCGACCCCTACACACACACCTTGTCTGACGCCCGCCACCTACGCGCTT 1014
Oy 161 GAlAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAspAsp 180
Db 1015 GGTGCGCATTAAGCTTTCTTCCGCTTACCGAGTCCAAAGCTAGAGCCTGCTGATGAT 1074
Oy 181 ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrGlnGlyValTyrAsnArgPhe 200
Db 1075 GTGGGCTGCTTCTTCTTCTTCAATTCAGGCTAGTATATACCAAGGGGATACACCGTTT 1134

Oy 201 AlaProTyr 203
Db 1135 GCTCCATAT 1143
RESULT 2
AF229057
LOCUS AF229057 2000 bp mRNA linear PRI 01-FEB-2002
DEFINITION Homo sapiens hexaribonucleotide binding protein 1 isoform gamma
ACCESSION (HNRNP1) mRNA, complete cds.
AF229057
VERSION AF229057.1 GI:18461366
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Chen, W. and Winkelman, J.C.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-2000) Internal Medicine/Hematology-Oncology,
University of Cincinnati College of Medicine, The Vontz Center for
Molecular Studies, 3125 Eden Avenue, Cincinnati, OH 45267-0508, USA
FEATURES
source location/Qualifiers
1..2000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p13.3"
1..2000
/gene="HNRNP1"
256..1512
/gene="HNRNP1"
/note="RNA-binding protein"
/codon_start=1
/product="hexaribonucleotide binding protein 1 isoform
gamma"
/protein_id="AA171904.1"
/db_xref="GI:18461367"
/translation="MLASQGVLLHPYGVPMIVPAAPYLGILQGNQAAAAADPTMAAP
YASQFAPQNGIPAEYTAAPHAPAEYTGQTVPEHTLNLNYPAAQTHSEOSPADTSA
QVSGATATQTDAAPTDGPOTOPSENNENKSORPLHVSNI PRPRPDRLRMEGOF
SKIDVEITIFERSRSKSGRTYFENSADADABEKLHCTVVEGRKIVNATAMTN
KTVNPTNGKLNIPVVGAVTSPEFVGVTLVLCQANGSSMTSAPSSLYTTSAMPF
PYPAATAAARGALRGRTVNTFPAALPPPIPAYGGVVQEPYVGNKLQGGY
AAYVAGTTPATAAAYSDSYGRVYAAPVHHALAPAPYGVGAMNAPFLTDATRSH
ADGVGVSSIQASITRYGQYRPAFY"
BASE COUNT 502 a 505 c 487 g 506 t
ORIGIN
Alignment Scores:
Pred. No.: 1.24e-69 Length: 2000
Score: 1063.00 Matches: 198
Percent Similarity: 98.52% Conservative: 2
Best Local Similarity: 97.54% Mismatches: 3
Query Match: 97.88% Indels: 0
DB: Gaps: 0
US-09-809-545A-2 (1-203) x AF229057 (1-2000)
Oy 1 MetThrAsnLysLysAlaValAsnProTyrThrAsnGlyTrpLysLeuAsnProValVal 20
Db 901 ATGACAAATAAAGAGCCGCTGACCCCTATACCAATGCTGGAATTAATCCAGTTG 960
Oy 21 GAlaValAlaTyrSerProAspPheTyrAlaGlyThrValLeuLeuCyseGlnAlaAsnGln 40
Db 961 GGTGCGATTAAGCTTCCGCAATTCATGACGACGCTCTTGTGCGACGCCAAACCG 1020
Oy 41 GAlGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
Db 1021 GAGGATCTTCCATGATGACGTCCCGCAGTTCACTTGTATATCTTCTGCAATGCGAGC 1080
Oy 61 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 80

Db 1081 TTCCCGATTCAGAGCCAGCCGCGGCGCTACCGAGGCGCCACCTCGAGGCGCC 1140
Qy 81 G1YArGThrValTYrAsnThrPheArgAlaAlaProProProProIlePheAlaTYr 100
Db 1141 GGTGGCACCGGTGTCACACCTTCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTAC 1200
Qy 101 G1Yg1YValTYrG1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Y 120
Db 1201 GCGCGAGTAGTATGATCAAGACCGTGTATGCAATTAATTGTCAGAGGCGGTATATCT 1260
Qy 121 AlArYrGThrValTYrAsnThrProAlaThrAlaAlaTYrSerAspSerTYrG1Y 140
Db 1261 GCATACCGCTAGCGCCAGCTTACCTCCACTGCGCTTACAGTACAGTACCGGA 1320
Qy 141 ArgValTYrAlaAlaAspProTYrHisThrLeuAlaProAlaProThrTYrG1YVal 160
Db 1321 CGATTATATGCTGCGACCCCTTACACAGCAGCTGCTCCAGCCCGCCACCTACGCGCTT 1380
Qy 161 G1YAlaMetAsnAlaPheAlaProLeuThraPalaTYrSerHisAlaAspAsp 180
Db 1381 GGTGCGCATGATGCTTTGCACTTTGACTGATGCGCAAGACTAGAGCCATGCTGATAT 1440
Qy 181 ValG1YValTYrG1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Y 200
Db 1441 GTGGGTCGCTTCTTCTTCATGTCAGGCTAGTATATACCGAGGCGGATACACCGTTT 1500
Qy 201 AlAProTYr 203
Db 1501 GCTCCATATC 1509
RESULT 3
AF191501 1586 bp mRNA linear ROD 01-MAR-2002
LOCUS Mus musculus hexaribonucleotide binding protein 1 (Hrbp1) mRNA,
DEFINITION complete cds.
ACCESSION AF191501 GI:19032413
VERSION AF191501.1 GI:19032413
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE 1. Mus musculus.
AUTHORS Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
TITLE 1 (bases 1 to 1586)
CHEN, W. and Winkelman, J.C.
JOURNAL Direct Submission
Submitted (01-OCT-1999) Internal Medicine/Hematology-Oncology,
University of Cincinnati College of Medicine, 231 Bethesda Ave.,
Cincinnati, OH 45267-0508, USA
FEATURES
source
1. 1586
/organism="Mus musculus"
/db_xref="taxon:10090"
/issue_type="skeletal muscle"
1. 1586
/gene="Hrbp1"
166..1419
/gene="Hrbp1"
/codon_start=1
/product="hexaribonucleotide binding protein 1"
/db_xref="GI:19032413"
/protein_id="AAL83425.1"
/translatability="TSLASQVLLHSYGVPMIYVPAAPYRGLMGNOEAAAAPDTMAOP
YASQAPSPONCIPAEYTAHPHAPAEYTGCTTYPDHNTLNYPTOTHSQSATISAO
TVSGTATCTDAAATPDGQPTQPSNTSKSQPKLHVSNIPIFRFPDRLQMGQCS
KILVEIIFNERSGKRGFTVPENSADDRAREKLGTVGGRKIENVNATAVWINK
KTVNPYINGWKLNPVGVVSPDYAGTVLLCQNOEGSSMYSGPSSLVYSAWGP
YPAASAAAYGALRGGRGTGYNTLRAAAPPIPAYGGVYDEPVYGNLTLCGVA
AYRAYOPTPATAAAYSDSGRYVADPYHTLAPAPTYGVGMNAPFLDTAKTRSHA
DDVGLVLSLQASLYRGYNFAPY"
BASE COUNT 407 a 458 c 374 g 347 t
ORIGIN
Alignment Scores:

Pred. No.: 1.34e-69 Length: 1586
Score: 1061.00 Matches: 198
Percent Similarity: 98.52% Conservative: 2
Best Local Similarity: 97.54% Mismatches: 3
Query Match: 97.70% Indels: 0
DB: 10 Gaps: 0
US-09-809-545a-2 (1-203) x AF191501 (1-1586)
Qy 1 MethThraenLYsAlaValAsnProTYrThraSngLYrTPLYsLeuAsnProValYal 20
Db 808 ATGACAAATTAAGAGAGCTGTACACCTTACACCAATGCTGGAATTAATCAAGTTGCT 867
Qy 21 G1YAlaValTYrSerProAspPheTYrAlaG1YThrValLeuLeuCYsG1AlaAsnG1N 40
Db 868 GCGCGGCTACAGCCCGGACTTCTATGACAGCAGCGGTGCTGTGTCAGAGCCAAACAG 927
Qy 41 G1UG1YSerSerMetTYrSerG1YProSerSerLeuValTYrThrSerAlaMetProG1Y 60
Db 928 GAGGAGTCTTCCATGATACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 987
Qy 61 PheProTYrProAlaAlaThrAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 80
Db 988 TTCCCATATCCGCGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1047
Qy 81 G1YArGThrValTYrAsnThrPheArgAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 100
Db 1048 GGTGCGACCGGCTACACACCTCAGGCGTACAGGCGCGCGCGCGCGCGCGCGCGCGCTAT 1107
Qy 101 G1Yg1YValTYrG1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Y 120
Db 1108 GCGGAGTAGTATGATCAAGACCGAGTGTATGCAATTAATTGCTACAGGCGGTATATCT 1167
Qy 121 AlArYrGThrValTYrAsnThrProAlaThrAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 140
Db 1168 GCGTACCGCTATGCTGCGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1227
Qy 141 ArgValTYrAlaAlaAspProTYrHisThrLeuAlaProAlaProThrTYrG1YVal 160
Db 1228 CGAGTTATGCTGCGACCCCTTACACAGCAGCTGCTCCAGCCCGCCACCTACGCGCTT 1287
Qy 161 G1YAlaMetAsnAlaPheAlaProLeuThraPalaTYrSerHisAlaAspAsp 180
Db 1288 GGTGCGCATGATGCTTTGCGCCCTTACAGCTGCAAGACTAGAGCCATGCTGATAT 1347
Qy 181 ValG1YValTYrG1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Yg1Y 200
Db 1348 GTGGGTCGCTTCTTCTTCATGTCAGGCTAGTATATACCGAGGCGGATACACCGTTT 1407
Qy 201 AlAProTYr 203
Db 1408 GCTCCATAT 1416
RESULT 4
AF107204 1363 bp mRNA linear ROD 23-JUN-2000
LOCUS Mus musculus ataxin 2-binding protein (A2bp) mRNA, complete cds.
ACCESSION AF107204
VERSION AF107204.1 GI:8671587
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE 1. Mus musculus.
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
TITLE 1 (bases 1 to 1363)
SHIBATA, H., HUYNH, D.P., VO, T.T. and PULST, S.-M.
JOURNAL A novel protein, A2BP, with RNA binding motif, binds to C-terminal
ataxin-2
Unpublished
2 (bases 1 to 1363)
REFERENCE
AUTHORS Shibata, H., Huynh, D.P., Vo, T.T. and Pulst, S.-M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1998) Division of Neurology and Neurogenetics

DB: 9 Gaps: 1

US-09-809-545a-2 (1-203) x AF109106 (1-3348)

Qy 1 MethraenlyblybAlaValAsnProtyrThrAsnGlyTyrPlybLeuAsnProValVal 20
 |||||
 Db 899 ATGCAAAATAAAGACCGTCAACCTTATCAAAATGGCTGAATTAATCCAGTTGTG 958

Qy 21 G1yAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuCyGlnAlaAsnGln 40
 |||||
 Db 959 GGTGACAGTCAAGTCCGGAATTCATATGACAGGACGCTCTGTTGTGCAAGCCACACAG 1018

Qy 41 GluGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
 |||||
 Db 1019 GAGGAGATCTTCATATGATACAGTGGCCCGGAGTTCACCTTGTATATCTTGTGCAATGCGACAGC 1078

Qy 61 PheProTyrProAlaAlaThrAlaAlaAlaAlaTyrArgGlyAlaHisLeuAlaGlyArg 80
 |||||
 Db 1079 TTCCCGATCCAGACGACCGCCGCGGCGCTTACCGAGGGCGCACCTCGACAGCCGC 1138

Qy 81 G1yArgThrValTyrAsnThrPheArgAlaAlaAlaProProProGlyProAlaTyr 100
 |||||
 Db 1139 GGTGACACCGTGTACACACCTTCAGGCGCGCGCGCCCGCCCGCCATCCCGCTAC 1198

Qy 101 G1yGlyValValTyrGlnGluProValTyrGlyAsnLybLeuGlnGlyTyrAla 120
 |||||
 Db 1199 GCGGAGATGATGATCAAGACCTGTGTATGCAATAATGCTCAGAGGTGTATGCT 1258

Qy 121 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaTyrSerAspSer----- 138
 |||||
 Db 1259 GCATATCCGCTACGCCACCTTACCCCTGCCACTGCCGTGCTACATGACAG- AAATCA 1317

Qy 139 -----TyrGlyArgVal 142
 |||||
 Db 1318 GTTGCTCTTCGTCAGACAGATGAATTTCTTTAAACCTTCGACGTACAGGAGGAGTT 1377

Qy 143 TyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGlyValGlyAla 162
 |||||
 Db 1378 TATGTCGCGACCCCTTACACACGACCTTGCTCAGGCCCGCCACCTACGCGGTGTGCTCC 1437

Qy 163 MetAsnAlaPheAlaProLeuThrAspAlaValTyrArgSerHisAlaAspAsyValGly 182
 |||||
 Db 1438 ATGATGCTTTTGACCTTTGATGATGCAAGACTAGAGGACATGATGATGTGGGT 1497

Qy 183 LeuValLeuSerSerLeuGlnAlaSerIleTyrGlnGlyGlyTyrAsnArgPheAlaPro 202
 |||||
 Db 1498 CTGCTCTTCTTCAATGACGCTATGATATACGAGGGGAGTACACCGTTTGTCTCA 1557

Qy 203 Tyr 203
 |||||
 Db 1558 TAC 1560

RESULT 6
 AB041596
 LOCUS Mus musculus brain cDNA, clone MNCD-3035, similar to Homo sapiens
 DEFINITION cDNA FLJ10105, clone HEMBA1003591.
 AB041596 2002 bp mRNA linear ROD 30-JUN-2000
 AB041596.1 GI:7670455
 VERSION 1.1 (full insert sequence)
 KEYWORDS Mus musculus (strain: C57BL) adult female cDNA to mRNA,
 SOURCE clone lib: Sugano mouse brain mncb clone: MNCD-3035.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (sites)
 AUTHORS Oseada, N., Kusuda, J., Tanuma, R., Ito, A., Hirata, M., Sugano, S. and
 Hashimoto, K.
 TITLE Isolation of full-length cDNA clones from mouse brain cDNA library
 made by oligo-capping method
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2002)
 AUTHORS Hashimoto, K., Oseada, N., Kusuda, J. and Sugano, S.

TITLE Direct Submission
 JOURNAL Submitted (12-APR-2000) Katsuyuki Hashimoto, National Institute of
 Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
 (E-mail: hashiku@nii.go.jp, URL: http://www.nii.go.jp/yoken/genebank/,
 Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)
 URL: http://www.nii.go.jp/yoken/genebank/
 COMMENT
 Lab Name: Sugano mouse brain mncb
 Lab host: TOP10
 Vector: pME18S-FL3
 1st strand cDNA was primed with an oligo (dT) primer
 [ATGTCGCTTTTCTTTTCTTTT]; double-stranded cDNA was synthesized
 using specific 5' and 3' primers and amplified by PCR. The PCR
 product was digested with SfiI and size selection was performed to
 exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
 into distinct DraII sites of pME18S-FL3. XhoI sites just outside
 the DraII sites can be used to isolate the cDNA insert. Library
 was constructed by Sugano et al. (University of Tokyo, Institute of
 Medical Science). Custom primer used for sequencing (5' end primer
 [CTTGTGCTTAAAGCTGCG]; 3' end primer
 [CGACCTGACCTCGACACAC]).
 A part of this sequence is reported in AU067167.
 FEATURES
 source
 Location/Qualifiers
 1..2002
 /organism="Mus musculus"
 /strain="C57BL"
 /db_xref="taxon:10090"
 /clone="MNCD-3035"
 /sex="female"
 /clone_lib="Sugano mouse brain mncb"
 /dev_stage="adult"
 637..1827
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAA95079.1"
 /db_xref="GI:7670456"
 /translation="MNCERQLRGNEAAAAPTMAQPVASQFAPFONGI PAEYAP
 HPHAPAYTGQTTPGHTLNLYPQTQHSQSDASQTSQTAQTDADA PDGQFQ
 TQSEMSKSGPKRLHVSNI PRFRPDRLDMGQSGKLLDVEI IENEGSKGFVY
 TFPNSADADBARLKHGTVEGRKIEVNNA TARYMTKTVNPYTNCKLNPVYAVY
 SPDPYACTVILCCANQEGSSMYSGPSLVTYTSAMPGPYPAATPAAAYRAHILGRGR
 TVNTERFAAPPPPIPAYGVSVVQDGYGVITGGVAVYVQPTPATPAAYSYSR
 VVADPHHTLAPFYGVGAMNNAFALTDKTRGHADVGLVLSLQAS1YRGYNR
 FAYP"
 BASE COUNT 518 a 589 c 516 g 379 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,69e-66 Length: 2002
 Score: 1020.50 Matches: 192
 Percent Similarity: 96.06% Conservative: 3
 Best Local Similarity: 94.58% Mismatches: 7
 Query Match: 93.97% Indels: 1
 DB: 10 Gaps: 1

US-09-809-545a-2 (1-203) x AB041596 (1-2002)

Qy 1 MethraenlyblybAlaValAsnProtyrThrAsnGlyTyrPlybLeuAsnProValVal 20
 |||||
 Db 1219 ATGCAAAATAAAGACCTGTAACCCCTTACACCAATGCTGGAATTAATCCAGTTGTG 1278

Qy 21 G1yAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuCyGlnAlaAsnGln 40
 |||||
 Db 1279 GCGCGGCTTACAGCCCGGACCTTATATGACAGGACGCTCTGTTGTGCAAGCCACACAG 1338

Qy 41 GluGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
 |||||
 Db 1339 GAGGAGATCTTCATATGATACAGTGGCCCGGAGTTCACCTTGTATATCTTGTGCAATGCGACAGC 1398

Qy 61 PheProTyrProAlaAlaThrAlaAlaAlaAlaTyrArgGlyAlaHisLeuAlaGlyArg 80
 |||||
 Db 1399 TTCCCGATCCGACGACCGCCGCGGCGCTTACCGAGGGCGCACCTCGACAGCCGC 1458

QY 81 G1yAgtThvAlaThrAsnThrPheAgaAlaAlaProProProProIleProAlaTy 100
 Db 1459 GGTCCACCGCTGTACAAACCTTACAGGCGTGCAGCGCCCGCCCAATCCCGGCTAT 1518
 QY 101 G1yG1yValValTyG1nG1uProValTyG1yAsnLysLeuLeng1nG1yG1yTyrAla 120
 Db 1519 GCGCGTGTGTGTACCAAGATGATTTATGGGACAGAC--ATTATATGCGATTATGCT 1575
 QY 121 A1ATyAgtTyAlaG1nProThrProAlaThraAlaAlaTySerAspSerTyG1y 140
 Db 1576 GCGTACCGCTATGCCAGCCACCTTGCACCTGCCGCGCTCACTACGATTAACGA 1635
 QY 141 ArgValTyTyrAlaAlaAspProTyRHisThrLeuAlaProAlaProThrTyG1yVal 160
 Db 1636 CGAGTTTATGTCGCCAGCCCTACACCACTGCTCCAGCCCGCCCACTACGCGCTT 1695
 QY 161 G1yAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAsp 180
 Db 1696 GGTCCATGAATGCTTTTCCGCTGACCGATGCGCAAGATGAGGCGCATGCTGATGAT 1755
 QY 181 ValG1yLeuValLeuSerSerLeuGlnAlaSerIleTyG1nG1yG1yTyrAsnArgPhe 200
 Db 1756 GTGGCTCTGCTTCTTCTTCATTGCGAGCTAGTATACCGAGGGGATACAACTTTT 1815
 QY 201 AlaProTyR 203
 Db 1816 GCTCCATAT 1824
 RESULT 7
 AK001027
 LOCUS Homo sapiens CDNA FLJ10165 f18, clone HEMBA1003591, weakly similar
 DEFINITION to CHLOROPLAST 28 KD RIBONUCLEOPROTEIN PRECURSOR.
 ACCESSION AK001027
 VERSION AK001027.1 GI:7022045
 KEYWORDS oligo capping, f18 (full insert sequence).
 SOURCE Homo sapiens embryo, 10 weeks whole embryo, mainly head CDNA to
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1
 AUTHORS Isogai,T., Oca,T., Hayaishi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
 Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
 Wagaatsuma,M., Hosokiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
 Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
 Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
 Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
 Nakamura,Y., Nagahara,K., Masuho,Y., Nimomiya,K. and Iwayanagi,T.
 TITLES NEDO human cDNA sequencing project
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1513)
 AUTHORS Isogai,T. and Otsuki,T.
 JOURNAL Direct Submission
 COMMENT Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,
 Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 (E-mail:genom@helix.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of
 International Trade and Industry of Japan; cDNA full insert
 sequencing; Research Association for Biotechnology; cDNA library
 construction, 5'- & 3'-end one pass sequencing and clone selection;
 Helix Research Institute (supported by Japan Key Technology Center
 etc.) and Department of Virology, Institute of Medical Science,
 University of Tokyo.
 FEATURES
 SOURCE
 1..1513
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HEMBA1003591"
 /cissue_type="whole embryo, mainly head"
 /clone_lib="HEMBA1"
 /dev_stage="embryo, 10 weeks"
 /note="cloning vector: pME18SFL3"

CDS
 221..1414
 /code="unamed protein product"
 /codon_start=1
 /protein_id="BA91472.1"
 /db_xref="GI:7022046"
 /translation="MNCERPOLRNOEAAAADPTMAOPYASQAPRPNQIPAEYTA
 PPHAPPEYTGQTVPEHTLNLPPAQTSEDSPADTSQATYSRTOQDDAAPDGGP
 QTPSEMENTSQPRRLVNSI PFRFRPDRLQFQGGKLLDVEIIFNERGKGF
 VTPENSADADPAREKLHGTVEGKI EYNNATARTMTKTVNPTNMGKLPVVGAV
 YSPEVAGTVLCOANOGSSMYSPASLSLVYTSAMPFPYPAATAAAYRGAHLRG
 RTVNTFRPAAPPPPIPAYGVVODGVADTYGVAAYVAOPTPATTAAYSDSYG
 RYVAADPRFHAAAPPTYGVMANMAFALTDAKRSHADVDVGLVLSQASIRGYN
 RPAFY"
 BASE COUNT 354 a 418 c 395 g 346 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 8e-66 Length: 1513
 Score: 1009.50 Matches: 190
 Percent Similarity: 95.57% Conservative: 4
 Best Local Similarity: 93.60% Mismatches: 8
 Query Match: 92.96% Indels: 1
 DB: 9 Gaps: 1
 US-09-809-545a-2 (1-203) x AK001027 (1-1513)
 QY 1 MetThrAsnLysLeuAlaValAsnProTyRHisGlyTyrLysLeuAsnProValVal 20
 Db 806 ATGACCAATAAAAAGCCGTCAACCTTATCAATGCTGGAAATGATTCAGTTGG 865
 QY 21 G1yAlaValTySerProAspPheTyRAlaG1yThrValLeuLeuCyGlnAlaGln 40
 Db 866 GGTGAGCTACAGTCCCAATTTATGACGACGCGTCTGTCGACGAGCCACG 925
 QY 41 G1yG1ySerSerMetTySerG1yProSerSerLeuValTyThSerAlaMetProGly 60
 Db 926 GAGGAGATCTTCATGACGTCGCCCACTTGTATATCTTCTCCATGCCAGGC 985
 QY 926 GAGGAGATCTTCATGACGTCGCCCACTTGTATATCTTCTCCATGCCAGGC 985
 QY 61 PheProTyRProAlaAlaThrAlaAlaAlaTyArgG1yAlaHisLeuArgGlyArg 80
 Db 986 TTCCGCTATCCAGCAGCCAGCCGCGCGCTACCGAGGGGCGCAGCTGCGAGGCGCC 1045
 QY 81 G1yAgtThvAlaThrAsnThrPheAgaAlaAlaProProProProIleProAlaTy 100
 Db 1046 GGTCCACCGCTGTACAAACCTTACAGGCGTGCAGCGCCCGCCCAATCCCGGCTAT 1105
 QY 101 G1yG1yValValTyG1nG1uProValTyG1yAsnLysLeuLeng1nG1yG1yTyrAla 120
 Db 1106 GCGCGTGTGTGTACCAAGATGATTTATGGGACAGAC--ATTATATGCGATTATGCT 1162
 QY 121 A1ATyAgtTyAlaG1nProThrProAlaThraAlaAlaTySerAspSerTyG1y 140
 Db 1163 GCATACCGCTATGCCAGCCCTACACCACTGCTCCAGCCCGCCCACTACGATTAACGA 1222
 QY 141 ArgValTyTyrAlaAlaAspProTyRHisThrLeuAlaProAlaProThrTyG1yVal 160
 Db 1223 CGAGTTTATGTCGCCAGCCCTACACCACTGCTCCAGCCCGCCCACTACGCGCTT 1282
 QY 161 G1yAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAsp 180
 Db 1283 GGTCCATGAATGCTTTTCCGCTGACCGATGCGCAAGATGAGGCGCATGCTGATGAT 1342
 QY 181 ValG1yLeuValLeuSerSerLeuGlnAlaSerIleTyG1nG1yG1yTyrAsnArgPhe 200
 Db 1343 GTGGCTCTGCTTCTTCTTCATTGCGAGCTAGTATACCGAGGGGATACAACTTTT 1402
 QY 201 AlaProTyR 203
 Db 1403 GCTCCATAT 1411
 RESULT 8
 AF107203
 LOCUS AF107203 2279 bp mRNA linear PRI 23-JUN-2000

DEFINITION Homo sapiens ataxin 2-binding protein (A2BP) mRNA, complete cds.
 ACCESSION AF107203
 VERSION AF107203.1 GI:8671585
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 2279)
 AUTHORS Shibata, H., Huynh, D.P., Vo, T.T. and Pulst, S.-M.
 TITLE A novel protein, A2BP, with RNA binding motif binds to C-terminal ataxin-2
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2279)
 AUTHORS Shibata, H., Huynh, D.P., Vo, T.T. and Pulst, S.-M.
 TITLE Direct Submission
 JOURNAL Submitted (16-NOV-1998) Division of Neurology and Neurogenetics Laboratory, Burns and Allen Research Institute, Cedars-Sinai Medical Center, UCLA School of Medicine, 8700 Beverly Blvd., Los Angeles, CA 90048, USA
 FEATURES
 source
 1..2279
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="Similar to Homo sapiens cDNA clone HSP38A20"
 1..2279
 /gene="A2BP"
 1047..2180
 /gene="A2BP"
 /codon_start=1
 /product="ataxin 2-binding protein"
 /protein_id="AAF78291.1"
 /translation="MAQPVASAPRONGIPAEYTAHPHPAPAEYTGTTPEHTLN LYPAPQHSQSPADTSQAQVSGRATDPAAPDGPQPSNTENKQPKRLHS NIPFERKDPDKQFGQFKILDEVEILFNERSGSGFEPFENSDADRKREKLGIV VEGRKLEVNNAATAVMTNKTVNPTVFMKLVNPGVAGVSPFEPYAGTVLLQANQES SWSAPSLVYTSAMPGPYPAPATAAAYGAMHFRGRVTYNTFRRAADPPPIPAYG GVVYODPFYADIDYGYAAVYAOPTPATYAAVSDSYGRVYAADPYHHALAPATYGV GANNAFAPLTDATKTRSHADVGLVSLQASIVRGYNRFAPY"
 1341..1358
 /gene="A2BP"
 /note="Region: RNA binding motif RNP-2"
 1452..1472
 /gene="A2BP"
 /note="Region: RNA binding motif RNP-1"
 BASE COUNT 497 a 712 c 636 g 434 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,26e-65 Length: 2279
 Score: 1009.50 Matches: 190
 Percent Similarity: 95.57% Conservative: 4
 Best Local Similarity: 93.60% Mismatches: 8
 Query Match: 92.96% Indels: 1
 Gaps: 1
 US-09-809-545a-2 (1-203) x AF107203 (1-2279)
 QY 1 MetThrAsnLysValAlaValAsnProTyrThrAsnGlyTyrPLeuLysLeuAsnProValAla 20
 DB 1572 ATGACAAATAAAAGACCGCTCAACCTTATACAAATGCTGGAATTGCAATCCAGTTG 1631
 QY 21 GAlaValAlaLysSerProAspPheTyrAlaGlyThrValLeuLeuCyGlnAlaAsnGln 40
 DB 1632 GGTCCAGCTCAAGTCCGCAATTCATGACAGCGGCTCTGTGTGCAGGCCAACACAG 1691
 QY 41 GAluGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
 DB 1692 GAGGATTTTCATATACAGTCCGCCAGTTCACTTGATATACCTTCGCAATGCCAGGC 1751
 QY 61 PheProTyrProAlaAlaThrAlaAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 80

DB 1752 TTCCCGATTCAGACGCCACCGCCGCCGCTACCGAGGGGGCCACTTCGAGGCCGC 1811
 QY 81 GAlaArgThrValTyrAsnThrPheArgAlaAlaAlaProProProPheAlaTyr 100
 DB 1812 GGTGCACCGGTGTACCAACCTTCAAGGCGCGCGGCCGCCGCCGATCCGGCTAC 1871
 QY 101 GAlaGlyValValTyrGlnGluProValTyrGlyAsnLysLeuLeuGlnGlyTyrAla 120
 DB 1872 GCGCGTGTGTTTACAGAGATGATTTATGTCACAGC--ATTATGGGTGTTATGCT 1928
 QY 121 AAlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaTyrSerAspSerTyrGly 140
 DB 1929 GCATACCGCTACGCCACCGCTACCCCTGCACCTGCTGCTACAGAGACAGTTACGA 1988
 QY 141 ArgValTyrAlaAlaAspProTyrHisThrLeuAlaProAlaProThrTyrGlyVal 160
 DB 1989 CGAGTTTATGCTGCGACCCCTACACACGACATTGCTCCAGCCGCCACTAGGGGTT 2048
 QY 161 GAlaAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAsp 180
 DB 2049 GGTCCATGAAATGCTTTGCACTTGACATGATCCAAAGACTAGAGCCATGCTGATGAT 2108
 QY 181 ValGlyLeuValLeuSerSerLeuGlnAlaSerLeuTyrGlnGlyGlyTyrAsnArgPhe 200
 DB 2109 GTGGGTCTGCTTCTTCTCATTCAGAGCTAGATATACGAGGGGATACACCGTTT 2168
 QY 201 AAlaProTyr 203
 DB 2169 GCTCATTC 2177
 RESULT 9
 AB060859
 LOCUS
 DEFINITION Macaca fascicularis brain cDNA clone: Qcra-11594, full insert sequence.
 ACCESSION AB060859.1 GI:13874510
 VERSION AB060859.1
 KEYWORDS oligo capping; fib (full insert sequence).
 SOURCE Macaca fascicularis adult male temporal lobe right cDNA to mRNA, clone 11b; macaque brain cDNA library Qcra clone: Qcra-11594.
 ORGANISM Macaca fascicularis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Macaca.
 REFERENCE 1 (sites)
 AUTHORS Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M., Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K.
 TITLE Isolation of full-length cDNA clones from macaque brain cDNA libraries
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1475)
 AUTHORS Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.
 TITLE Direct Submission
 JOURNAL Submitted (27-APR-2001) Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
 (E-mail: khashi@nih.go.jp, URL: http://www.nih.go.jp/yoken/genbank/, Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)
 Lab Host: TOP10
 Vector: pME18S-FL3 (Acc. No. AB009864)
 R. Site1: DraIII (CACTGTTGG)
 R. Site2: DraIII (CACCATGG)
 Description: 1st strand cDNA was primed with an oligo(dT) primer [ATGTCGCTTTTCTTTTCTTTT]; double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18S-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by oligo-capping method (Sugano et al., Institute of Medical Science, University of Tokyo).
 Custom primer used for sequencing

(5' end primer (CTTCTGCTTAAAGCTGCG) ;
3' end primer (CGACCTGACGCTGACAC)).
Location/Qualifiers
1. 1475
/organism="Macaca fascicularis"
/db_xref="taxon:9641"
/clone="Qtra-11594"
/sex="male"
/tissue_type="temporal lobe right"
/clone_lib="macaque brain cDNA library Qtra"
/dev_stage="adult"
90. 1220
/codon_start=1
/product="hypothetical protein"
/protein_id="BAB46877.1"
/db_xref="GI:13874511"

CDS

/translation="MEKGRMNOGNOEAAAPPTMAOPYASQAFAPONGIPAEVT
APHPAPAEVGTTPPEHTLNLIPPAOTHESESPADTNAGTSTATQTDAAPTDG
QPTOPSENTENKSPKRLHVSNIPEFRPDRLQMFQFGKILDELIFENRSGKF
GFTVFENADARAEKLGIVEGRIEVMNATKTVIPYNGKLVNPG
AVGPEFVAGTVLCOANOEGSSMSYSPSSLYTSAMGPPYPATATAAAYGAILRG
RGRTVYNTFRRAAAPPPPIPAYGVVYODGPGADIVGGYALYRVAYQTPATAAAYSDR
NQPVPVADDEISNTSATVTDEFMLPTPTTHLQPPITALYP"
BASE COUNT 389 a 414 c 363 g 309 t
ORIGIN

Alignment Scores:

Pred. No.: 8,92e-63 Length: 1475
Score: 968.00 Matches: 189
Percent Similarity: 87.33% Conservative: 4
Best Local Similarity: 85.52% Mismatches: 9
Query Match: 89.13% Indels: 20
DB: 9 Gaps: 2

US-09-809-545A-2 (1-203) x AB060859 (1-1475)

QY 1 MetThrAsnLysLysAlaValAsnProTyrThrAsnGlyTTrpLysLeuAsnProValVal 20
Db 681 ATGACAAATATAAAGACCGTCAACCTTATACAAATGCGTGAATTTGAATCCAGTTGTG 740
QY 21 GtAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuGlyGlnAlaAsnGln 40
Db 741 GGTGGGCTTACAGTCCGGAATTCTATGACGACGCGCTCTGTGTGCCAGGCCAACCG 800
QY 41 GtGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
Db 801 GAGGAGCTTCCATGATGACGTGCCCGCCAGTTCCTGTATATCTTCGCAATGCCAGGC 860
QY 61 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 80
Db 861 TTCGGGTATCCAGCACCGCCGCGCGCGCTTACCGAGGGCGCACCTGCGAGGCCCGC 920
QY 81 GtAlaGlyThrValTyrAsnThrPheArgAlaAlaAlaProProProProIleProAlaTyr 100
Db 921 GCCCGGACCGGTGACAAACCTTCAGGGCGCGAGCCCCCGCGCATCCCGGCTAT 980
QY 101 GtAlaValValTyrGlnGluProValTyrGlyAsnLysLeuLeuGlnGlyTyrAla 120
Db 981 GCGGCTGTTTTCACGAGATGATTTTATGTGTGACAGC---ATTATGTGTGTTATGCT 1037
QY 121 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaTyrSerAspSer----- 138
Db 1038 GCATACCGGTACGCCAGCTACCCCTGCGACGGCGCGCTCTTACGATGACAG-AAATCA 1096
QY 139 -----TyrGlyArgVal 142
Db 1097 GTTCGCTTCTGTCAGCAGATGAATTTCTTGTAAACCTCTGACGTTACGACAGATT 1156
QY 143 TrrAlaAlaAspProTyrHisHisIleThrLeuAlaProAlaProThrTyrGlyValGlyAla 162
Db 1157 TATGTGCCGACCCCTTACCAACGACCTTGCTCCAGCCCCCCTTACGCGCTGTGGTCC 1216
QY 163 MetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAspAspValGly 182

Db 1217 ATGAATGCTTTTGACCTTTGACGTATGCCAAGACTTACGATCTGATGATGTGGT 1276
QY 183 LeuValLeuSerSerLeuGlnAlaSerIleTyrGlnGlyTyrAsnArgPheAlaPro 202
Db 1277 CTCGCTTCTTCTTCAATTCGAGCTAGTATATACGAGGGGGATACACCGTTTCTCCA 1336
QY 203 Tyr 203
Db 1337 TAC 1339

RESULT 10

ARI34676 2372 bp DNA linear PAT 16-MAY-2001
LOCUS ARI34676
DEFINITION Sequence 1 from patent US 6194171.
ACCESSION ARI34676
VERSION ARI34676.1 GI:14123581
KEYWORDS
SOURCE
ORGANISM
Unknown.
REFERENCE
1 (bases 1 to 2372)
AUTHORS Pulst, S.M. and Shibata, H.
TITLE Nucleic acids encoding ataxin-2 binding proteins
JOURNAL Patient: US 6194171-A 1 27-FEB-2001;
FEATURES
source location/Qualifiers
1. 2372
/organism="unknown"

BASE COUNT

548 a 726 c 646 g 452 t
ORIGIN

Alignment Scores:

Pred. No.: 1.38e-61 Length: 2372
Score: 955.00 Matches: 189
Percent Similarity: 86.94% Conservative: 4
Best Local Similarity: 85.14% Mismatches: 9
Query Match: 87.94% Indels: 21
DB: 6 Gaps: 2

US-09-809-545A-2 (1-203) x ARI34676 (1-2372)

QY 1 MetThrAsnLysLysAlaValAsnProTyrThrAsnGlyTTrpLysLeuAsnProValVal 20
Db 1572 ATGACAAATATAAAGACCGTCAACCTTATACAAATGCGTGAATTTGAATCCAGTTGTG 1631
QY 21 GtAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuGlyGlnAlaAsnGln 40
Db 1632 GGTGAGTCTACAGTCCGGAATTCTATGACGACGCGCTCTGTGTGCCAGGCCAACCG 1691
QY 41 GtGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
Db 1692 GAGGAGCTTCCATGATGACGTGCCCGCCAGTTCCTGTATATCTTCGCAATGCCAGGC 1751
QY 61 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 80
Db 1752 TTCGGGTATCCAGCACCGCCGCGCGCGCTTACCGAGGGCGCACCTGCGAGGCCCGC 1811
QY 81 GtAlaGlyThrValTyrAsnThrPheArgAlaAlaAlaProProProProIleProAlaTyr 100
Db 1812 GGTGCGACCGGTGACAAACCTTCAGGGCGCGAGCCCCCGCGCATCCCGGCTATC 1871
QY 101 GtAlaValValTyrGlnGluProValTyrGlyAsnLysLeuLeuGlnGlyTyrAla 120
Db 1872 GCGGCTGTTTTCACGAGATGATTTTATGTGTGACAGC---ATTATGTGTGTTATGCT 1928
QY 120 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaTyrSerAspSer----- 138
Db 1929 TGCATACCGCTACGCCAGCTTACCCCTGCGACGGCGCTTACGATGACAG-AAATC 1987
QY 139 -----TyrGlyArgVal 142
Db 1988 AGTTGCTTCTGTCAGCAGATGAATTTCTTGTAAACCTCTGACGATTACGAGAGAGT 2047

QY 142 1TyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGlyValGlyAl 162
 Db 2048 TTATGCTGCCACCCCTACACACGACCTGCTCCAGCCCCCACTACGGCGTTGGTGC 2107

QY 162 aMeAspAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAspPheValG 182
 Db 2108 CATAAAGCTTTTGACCTTGACGTGATGCAAGCTGAGGACCATGCTGATGATGGG 2167

QY 182 YLeuValLeuSerSerLeuGlnAlaSerIleTyrGlnGlnGlyTyrAsnArgPheAlaPr 202
 Db 2168 TCTGCTTCTTCTTCATTCATTCGACGCTAGATATACCGAGGGGATACACCGTTTGCTCC 2227

QY 202 cTyr 203
 Db 2228 ATAC 2231

RESULT 11
 AF094849 1547 bp mRNA linear PRI 01-MAR-2002
 LOCUS Homo sapiens hexaribonucleotide binding protein 1 isoform alpha
 DEFINITION (HRNB1) mRNA, complete cds.
 ACCESSION AF094849
 VERSION AF094849.1 GI:19032365
 KEYWORDS Homo sapiens.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1547)
 Chen, W., Chu, Z.-L., Blough, R. I., Liu, L., Hoppes, B. and Winkelman, J. C.
 Molecular cloning and chromosomal localization of a human brain, heart and skeletal muscle specific RNA binding protein gene homologous to fox-1 in Caenorhabditis elegans

TITLE Unpublished
 2 (bases 1 to 1547)
 Chen, W., Chu, Z.-L., Blough, R. I., Liu, L., Hoppes, B. and Winkelman, J. C.
 Direct Submission
 Submitted (24-SEP-1998) Internal Medicine/Hematology-Oncology, University of Cincinnati College of Medicine, 231 Bethesda Ave., Cincinnati, OH 45267-0508, USA
 Location/Qualifiers
 1. 1547
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /map="16p13.3"
 1. 1547
 /gene="HRNB1"
 254. 1432
 /gene="HRNB1"
 /note="similar to Caenorhabditis elegans fox-1"
 /codon_start=1
 /product="hexaribonucleotide binding protein 1 isoform alpha"
 /protein_id="AAL3405.1"
 /db_xref="GI:19032365"
 /translation="MLASQGVLLHPYGVPMIVPAAPYLPGILQNGEAAAPDTMAOP
 VASQAPQNGIPNGIAPAEYAPHPAPPEYGTQTVBEHTLNLVPAQTHSESPADTSA
 QTVSGTATQDDADPTDGPOTQPSSENTENKSPKRLHVSNIIPRFDPDLROMQCF
 GTILDEVEILFNERSKGFVTFENSADADARERKLGTVESRKIEVNNAIARMTN
 KKTINPTNGMKINPVGAVSPERFVATGIVLCOANGSGMSAPSLVYTSMPGF
 PYPAATAAAYRGHLKRGRTVYNTRRAAAPPPPIATGAGVYQEEVYNGKLQGGY
 AAAYRAQPTPATAAAYSDSYGRVYAADPYHHLAPATLVGVMAASTYRGYNFAPY"

BASE COUNT 380 a 448 c 374 g 345 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.27e-58 Length: 1547
 Score: 912.00 Matches: 172
 Percent Similarity: 85.71% Conservative: 2

Best Local Similarity: 84.73% Mismatches: 3
 Query Match: 83.98% Indels: 26
 DB: 9 Gaps: 1

US-09-809-545A-2 (1-203) x AF094849 (1-1547)

QY 1 MetThrAsnLysLysAlaValAlaAsnProTyrThrAsnGlyTyrLysLeuAsnProValAl 20
 Db 899 ATGACAAATAAAGAACCCGTCAACCTTATACAAATGGCTGGAATGAATCCAGTTGTG 958

QY 21 GlyAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLysGlnAlaAsnGln 40
 Db 959 GGTGACGTACAGTCCCGAATTCTATGACAGCACGGTCTGTGTGTCAGGCGCAACAG 1018

QY 41 GluLysSerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
 Db 1019 GAGGATCTTCATGATGACAGGCGCCCACTTACATTGATATATCTTCGCAATGCGAGCG 1078

QY 61 PheProTyrProAlaAlaThrAlaAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 80
 Db 1079 TTCCCGTATCCAGACAGCACCGCGCGCCCTACCGAGGCGCACCTCGAGGCGCG 1138

QY 81 GlyArgThrValTyrAsnThrPheArgAlaAlaAlaProProProProIleProAlaTyr 100
 Db 1139 GGTGCACCGGTGACACACCTTCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTAC 1198

QY 101 GlyGlyValValTyrGlnGlnProValTyrGlyLysAsnLysLeuGlnGlyTyrAla 120
 Db 1199 GCGGAGTACGTATCAAGACCTGTGTATGCAATAAATGCTGCGAGGTGTTATCT 1258

QY 121 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaTyrSerAspSerTyrGly 140
 Db 1259 GCATACCGCTACGCCACCGCTACCCCTGCGCGCGCGCTACGAGTACGAGTACGAG 1318

QY 141 ArgValTyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGlyVal 160
 Db 1319 CGAGTTATGTGCGGACCCCTACACACACCTTGTCTCAGGCCCGCACCTACGCGCTT 1378

QY 161 GlyAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAspAsp 180
 Db 1379 GGTGCGCATG-----GCTAGTATATCCAGAGGGGATACACCGTTT 1420

QY 181 ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrGlnGlyTyrAsnArgPhe 200
 Db 1388 -----GCTAGTATATCCAGAGGGGATACACCGTTT 1420

QY 201 AlaProTyr 203
 Db 1421 GCTCCATAC 1429

RESULT 12
 AY072786 1393 bp mRNA linear PRI 22-MAR-2002
 LOCUS Homo sapiens RNA binding motif protein 9 mRNA, complete cds.
 DEFINITION Homo sapiens RNA binding motif protein 9 mRNA, complete cds.
 ACCESSION AY072786
 VERSION AY072786.1 GI:19584571
 KEYWORDS Homo sapiens.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1393)
 Norris, J. D., Fan, D., Sherck, A. and McDonnell, D. P.
 A negative coregulator for the human ER
 Mol. Endocrinol. 16 (3), 459-466. (2002)

TITLE A negative coregulator for the human ER
 JOURNAL Mol. Endocrinol. 16 (3), 459-466. (2002)
 MEDLINE 21864496
 PUBMED 11875103
 REFERENCE 2 (bases 1 to 1393)
 Norris, J. D., Fan, D. and McDonnell, D. P.
 Direct Submission
 Submitted (10-JUN-2002) Pharmacology and Cancer Biology, Duke University Medical Center, Research Drive, USMC Building, Rm C264, Durham, NC 27710, USA

```

FEATURES
    source
        Location/Qualifiers
            1..1393
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="22"
                /map="22q13.1"
                /cell_line="HeLa"
            106..1278
                /note="RBM9; splice variant; RPA; repressor of tamoxifen
                transcriptional activity"
                /codon_start=1
                /product="RNA binding motif protein 9"
                /protein_id="AA067150.1"
                /db_xref="GI:19584572"
                /translation="MONEPLTPGHHGPARDSDGNOEPTTPDAMVOPFTTIPPPPP
                ONGIPTEYGVPHTDYAGOTGSHNLTLYGSOAHGEOSNSPSTONGSLTTEGCAQTD
                GQSGTOSSENSSEKSTPKRLHVSNI PRPRPDRLROMFGGKILDEIITFENRGSK
                GGFRTFENSADADKARKLHGTVEGRKIVNNATARVMTNKKVTPYANGKSLSPVYAGV
                VGVAYGPELVLAASFDADVSLGNDAAVPLSGRGGINITITPLISLPLVGFPRPTAAT
                AAAPFGAHLRGGRVGVAVRAVPTALPAYGVVYQDGFAGADLYGVAAVRYAQPA
                TATATAAAAAAASDYGVRVYADPYHALAPAAVGVAVASLYRGYSRFABY"
            442..711
                /note="Region: RNA recognition motif"
                /misc_feature
                    /note="Region: RNA recognition motif"
BASE COUNT      361 a      376 c      379 g      277 t
ORIGIN
Alignment Scores:
Pred. No.:      3.25e-30      Length:      1393
Score:          526.00      Matches:      116
Percent Similarity: 59.55%      Conservative: 15
Best Local Similarity: 52.73%      Mismatches: 43
Query Match:     48.43%      Indels:      46
DB:              9      Gaps:      8

US-09-809-545A-2 (1-203) x AY072786 (1-1393)
QY      1 MetTrrAenLyeLyeAlaValaenProTyrThrAenGlyTrrLyeLyeLyaenProValaVal 20
Db      703 ATGACCAATTAAGACAGATGCTACACCATATGCAAATGCTTGGAATTAAGCCAGTAGTT 762
QY      21 GtYAlaValaTyrSerProAenPheTyrAlaGlyThrValLeuLeuCy6GlnAlaAen--- 39
Db      763 GGAGCGTATATGCTCCGAGATTATGACGATCCAGCTTTCACAGCATGTGTCCCTA 822
QY      40 ---GlnGlySerSerMetTyrSerGlyProSerSer-----Leu 52
Db      823 GGCATGATGACGACGACATGCCCTATCAGAGAAGGGGGATCAACAATTACATCTCTTA 882
QY      53 ValTyrThrSerAlaMetProGlyPheProTyrPro---AlaAlaThrAlaAlaAla 71
Db      883 ATCAGTCTCCCTTATGTTCTCTGCTTCCCTTACCTTACCTGACCCACGCGGCGCGCT 942
QY      72 TrrYrrGtAlaAlaHisLeuAenGtYrrGtYrrGtYrrValTrrAenThrPheAtrAlaAla 91
Db      943 TTCAGAGGAGCCCATTTGAGGGGAGAGGGGAGCATATGCTGCACTCCA---GCG 999
QY      92 AlaProProProProIleProAlaTyrGtYrrValTrrGtYrrGtYrrGtYrrGtYrrGtYrr 111
Db      1000 GtACCTCCCAACAGCCATCCCGCTTATCCAGGTGTGTTTACAGAGAGGATTTTACGGT 1059
QY      112 AenLyeLyeLyeGtYrrGtYrrAlaAlaTrrYrrGtYrrAlaGtYrrAlaGtYrrProAlaThr 131
Db      1060 GCTGAC---CTGTATGTGTGATATGACGCTTACAGATATGACAGCTGCTACTGCAACC 1116
QY      132 -----AlaAlaAlaTrrSerAenSerTrrGtYrrGtYrrGtYrrGtYrrGtYrrGtYrr 143
Db      1117 GCAGCCACCGCTGCTGACCGCTGACGCGCTTACAGTACGCGTTATGACAGGGGTGTAC 1176
QY      144 AlaAlaAenProTrrHisThrLeuAlaProAlaProThrTrrGtYrrGtYrrGtYrrGtYrrGtYrr 163
Db      1177 ACAGCGGAGCCCTTAC---CATGCCCTTGGCCCTGCGCGCTAGCTATGAGAGTGGCGCTGTG 1233
QY      164 AenAlaPheAlaPheLeuThrAenAlaLyeThrAenSerHisAlaAenAenValGtYrrLeu 183

```

```

Db      1233 -----
QY      184 ValLeuSerSerLeuGlnAlaSerIleTrrGtYrrGtYrrGtYrrGtYrrGtYrrGtYrr 203
Db      1234 -----GCGAGTTTATACGAGGTTGCTACAGCCGATTTGCCCTTAC 1275

RESULT 13
LOCUS      BC025281
DEFINITION Homo sapiens, RNA binding motif protein 9, clone MGC:39192
ACCESSION BC025281
VERSION    BC025281.1
KEYWORDS   GI:19264129
SOURCE     MGC.
ORGANISM   Homo sapiens.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 1555)
REFERENCE   Strausberg, R.
            Direct Submission
            Submitted (05-MAR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
            NIH-MGC Project URL: http://mgc.nci.nih.gov
REMARK     Contact: MGC help desk
            Email: cga@bbs-remail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Rubin Laboratory
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Sueanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Teal, Natassja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 42 Row: d Column: 8
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 7657503.

FEATURES
    source
        Location/Qualifiers
            1..1555
                /organism="Homo sapiens"
                /db_xref="LocusID:23543"
                /db_xref="taxon:9606"
                /clone="MGC:39192 IMAGE:4555865"
                /tissue_type="uterus, leiomyosarcoma"
                /clone_id="NIH_MGC_46"
                /lab_host="DH10B-R"
                /note="Vector: pOTB7"
            189..1331
                /codon_start=1
                /product="RNA binding motif protein 9"
                /protein_id="AAH25281.1"
                /db_xref="GI:19264130"
                /translation="MEKKMYTQGNQEPPTTPDAMVOPFTTIPPPPPONGIPTEYGV
                PHTDYAGOTGSHNLTLYGSOAHGEOSNSPSTONGSLTTEGCAQTDGQSGTOSSENSSEK
                STPKRLHVSNI PRPRPDRLROMFGGKILDEIITFENRGSKGGFRTFENSADADKARKLHGT
                VEGRKIVNNATARVMTNKKVTPYANGKSLSPVYAGVAVGVAYGPELVLAASFDADVSLGND
                AAAPFGAHLRGGRVGVAVRAVPTALPAYGVVYQDGFAGADLYGVAAVRYAQPAATATAAAAA
                AASDYGVRVYADPYHALAPAAVGVAVASLYRGYSRFABY"
CDS
BASE COUNT      469 a      371 c      383 g      332 t
ORIGIN

```

Alignment Scores:

Pred. No.: 3,68e-30 Length: 1555
 Score: 526.00 Matches: 116
 Percent Similarity: 59.55% Conservative: 15
 Best Local Similarity: 52.73% Mismatches: 43
 Query Match: 48.43% Indels: 46
 DB: 9 Gaps: 8

US-09-809-545A-2 (1-203) x BC025281 (1-1555)

```

Qy 1 MetThrAsnLysValAlaAsnProTyrThrAsnGlyTTPlysLeuAsnProValVal 20
Db 756 ATGACCAATAAGAAATGATGTCACACCATATGCAATATGTTGAATTAAGCCCACTAGTT 815
Qy 21 GlyAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuCysGlnAlaAsn--- 39
Db 816 GGAGCTGTAATGCTCCGAGATTATATGACATTCAGCTTCAAGCAGATGTCTCCTA 875
Qy 40 --GlnGlySerSerMetTyrSerGlyProSerSer-----Leu 52
Db 876 GGCAATATGACGACGTCGCTCATCAGGAAGAGGGGATCAACACTTACATTCCTTTA 935
Qy 53 ValTyrThrSerAlaMetProGlyPheProTyrPro--AlaAlaThrAlaAlaAla 71
Db 936 ATCAGTCTCCCTTATGTTCTGCTTCCCTTACCTACTGACGACGACGACGACGCT 995
Qy 72 TyrArgGlyAlaHisLeuArgGlyValArgThrValTyrAsnThrPheArgAlaAla 91
Db 996 TTCAGAGAGCCCATTTAAGGGGAGGCGGACAGATATGATGTGACGTCGCA--GCG 1052
Qy 92 AlaProProProPheProAlaProAlaTyrGlyValValTyrGlnGluProValTyrGly 111
Db 1053 GTACTTCCAAAGCAGCATCCCGCCCTATCCAGGTGTGTTTACAGAGGAGGATTTACGCT 1112
Qy 112 AsnLysLeuLeuGlnGlyTyrAlaAlaTyrArgTyrAlaGlnProThrProAlaThr 131
Db 1113 GCTGAC--CTCTATGTGTGATATGACAGCTTACAGATATGACAGACCTGCTCAAC 1169
Qy 132 -----AlaAlaAlaTyrSerAspSerTyrGlyArgValTyr 143
Db 1170 GCAGCCACCGCTGTCGACGCGCTGACGCTTACAGTACGATATGCGAGGATGAC 1229
Qy 144 AlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGlyValGlyAlaMet 163
Db 1230 ACAGCCGACCCCTAC--CATGCCCTTGCCCTGCGCTAGCTATGAGCTTGCCCTG 1286
Qy 164 AsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAspAspValGlyLeu 183
Db 1286 ----- 1286
Qy 184 ValLeuSerSerLeuGlnAlaSerIleTyrGlnGlyTyrAsnArgPheAlaProTyr 203
Db 1287 -----GCGAGTTTATACCGAGGTGCTACAGCGCATTTGCCCTTAC 1328

```

RESULT 14
 AF229058 1925 bp mRNA linear PRI 01-FEB-2002
 DEFINITION Homo sapiens hexaribonucleotide binding protein 2 (HRNP2) mRNA,
 complete cds.
 ACCESSION AF229058
 VERSION AF229058.1 GI:18461368
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 1925)
 TITLE Chen, W. and Winkelman, J.C.
 JOURNAL Direct Submission
 Submitted (28-JUN-2000) Internal Medicine/Hematology-Oncology,
 University of Cincinnati College of Medicine, The Vontz Center for
 Molecular Studies, 3125 Eden Avenue, Cincinnati, OH 45267-0508, USA

FEATURES

source

Location/Qualifiers
 1..1925
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="22"
 1..1925
 /gene="HRNP2"
 171..1313
 /gene="HRNP2"
 /note="RNA-binding protein; alternatively spliced RBM9"
 /codon_start=1
 /product="hexaribonucleotide binding protein 2"
 /protein_id="AAL71905.1"
 /db_xref="GI:18461369"
 /translation="MEKKKVVTOGNOEPTTTPDPAVQFTTTPPPPPONGIPREYGV
 PHODVAGOTGEHNLTLVYSTOAHGEOSNSPTSONSLTTEGAOTDGOOSOTOSE
 NSSSKSPKRIHYSNIPFRPDIDRMPOQFQGLIDLEIIFNRSKSGCFYFENS
 ADDDRAREKLGTVBGRKTEVNNAATKMTNKMTTPYANGKLSYVGAVPELY
 AASSFDQVSLGNDAAVPLSGRGINTYIPILSLPLVPGFPPTAATTAAPRAHLR
 GRGRTVGAARAIVPTAIPAYPGVYODGEYADLYXGYAARYAOCATATATMAAA
 AAAAYSDGYGRVYTAADPYHALAPASVGVAVASLYRGYSRFAPY"

CDS

BASE COUNT

562 a 436 c 467 g 460 t
 ORIGIN

Alignment Scores:

Pred. No.: 4.67e-30 Length: 1925
 Score: 526.00 Matches: 116
 Percent Similarity: 59.55% Conservative: 15
 Best Local Similarity: 52.73% Mismatches: 43
 Query Match: 48.43% Indels: 46
 DB: 9 Gaps: 8

US-09-809-545A-2 (1-203) x AF229058 (1-1925)

```

Qy 1 MetThrAsnLysValAlaAsnProTyrThrAsnGlyTTPlysLeuAsnProValVal 20
Db 738 ATGACCAATAAGAAATGATGTCACACCATATGCAATATGTTGAATTAAGCCCACTAGTT 797
Qy 21 GlyAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuCysGlnAlaAsn--- 39
Db 798 GGAGCTGTAATGCTCCGAGATTATATGACATTCAGCTTCAAGCAGATGTCTCCTA 857
Qy 40 --GlnGlySerSerMetTyrSerGlyProSerSer-----Leu 52
Db 858 GGCAATATGACGACGTCGCTCATCAGGAAGAGGGGATCAACACTTACATTCCTTTA 917
Qy 858 GGCAATATGACGACGTCGCTCATCAGGAAGAGGGGATCAACACTTACATTCCTTTA 917
Qy 53 ValTyrThrSerAlaMetProGlyPheProTyrPro--AlaAlaThrAlaAlaAla 71
Db 918 ATCAGTCTCCCTTATGTTCTGCTTCCCTTACCTACTGACGACGACGACGCT 977
Qy 72 TyrArgGlyAlaHisLeuArgGlyValArgThrValTyrAsnThrPheArgAlaAla 91
Db 978 TTCAGAGAGCCCATTTAAGGGGAGGCGGACAGATATGCTGACGTCGCA--GCG 1034
Qy 92 AlaProProProPheProAlaProAlaTyrGlyValValTyrGlnGluProValTyrGly 111
Db 1035 GTACTTCCAAAGCAGCATCCCGCCCTATCCAGGTGTGTTTACAGAGGAGGATTTACGCT 1094
Qy 112 AsnLysLeuLeuGlnGlyTyrAlaAlaTyrArgTyrAlaGlnProThrProAlaThr 131
Db 1095 GCTGAC--CTCTATGTGTGATATGACAGCTTACAGATATGACAGCCTGCTACTGCAAC 1151
Qy 132 -----AlaAlaAlaTyrSerAspSerTyrGlyArgValTyr 143
Db 1152 GCAGCCACCGCTGTCGACGCGCTGACGCTTACAGTACGATATGCGAGGATGAC 1211
Qy 144 AlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGlyValGlyAlaMet 163
Db 1212 ACAGCCGACCCCTAC--CATGCCCTTGCCCTGCGCTAGCTATGAGCTTGCCCTG 1268
Qy 164 AsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAspAspValGlyLeu 183
Db 1268 ----- 1268

```

```

QY 184 ValLeuSerLeuGlnAlaSerIleTyrgInglyGlyTyrAsnArgPheAlaProTyr 203
Db 1269 -----GCGAGTTATACCGAGGTGCTACAGCCGATTGGCCCCCTAC 1310

RESULT 15
LOCUS AF387322 1134 bp mRNA linear ROD 20-JUN-2001
DEFINITION Mus musculus putative RNA-binding protein fah mRNA, complete cds.
ACCESSION AF387322
VERSION AF387322.1 GI:14495355
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Mus musculus; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1134)
AUTHORS Lieberman, A.P., Friedlich, D.L., Harrison, G., Howell, B.W.,
Jordan, C.L., Breedlove, S.M. and Fischbeck, K.H.
JOURNAL Androgens regulate the mammalian homologues of invertebrate sex
BIOCHEM. Biophys. Res. Commun. 282 (2), 499-506 (2001)
MEDLINE 21294761
JOURNAL 11401487
PUBMED 2 (bases 1 to 1134)
REFERENCE Lieberman, A.P. and Fischbeck, K.H.
AUTHORS Direct Submission
JOURNAL Submitted (25-MAY-2001) Neurogenetics Branch, National Institutes
of Health, Building 10, Room 3B11, Bethesda, MD 20892, USA
FEATURES
Source Location/Qualifiers
1..1134
/organism="Mus musculus"
/db_xref="taxon:10090"
1..1134
/note="similar to Caenorhabditis elegans fox-1; contains
RBM domain"
/codon_start=1
/product="putative RNA-binding protein fah"
/protein_id="AAK64287.1"
/db_xref="GI:14495355"
/translation="MEKKMVTGNOEPTTPDAMVOPPTTIPPPPPONGIPTEYGV
PHQDYAGQTSRHNLTLYGSTPHREGSSNSPNSONGISTOTEGSAQDQGSQSS
ENSESKSTPKRLHVSNIPIPRFSDPDLRFQFGKILDELIFNERSKGFVTFEN
SADADRAEKLGHTVEGRIEVENNAIATVMTNKKMVTPIYANGWKLSPVGAIVPEL
YAASFQADVLSNEAAVPLSGRGINTYPIIIPFPYPTATTAARFGAHLRGRG
RTVGVAVRVPPTAI PAYGVYODGFGADLYGGVAAVRYAATATATATATATAA
AYSDGVRVYTDPRYHALAPASVGVAVASLYRGYSRFAY"

BASE COUNT 301 a 312 c 293 g 228 t
ORIGIN

Alignment Scores:
Pred. No.: 3,06e-30 Length: 1134
Score: 525.00 Matches: 116
Percent Similarity: 60.198 Conservative: 14
Best Local Similarity: 53.708 Mismatches: 44
Query Match: 48.344 Indels: 42
DB: 10 Gaps: 8

US-09-809-545a-2 (1-203) x AF387322 (1-1134)
QY 1 MetThrAsnLysAlaValaValaProTyrThrAsnGlyTyrPlyLeuAlaProVal 20
Db 571 ATGACCAACAGAAAGTGTGACGCATATGCAAAATGGCTGAAGTTAAGCCAGTAGTT 630
QY 21 GlyAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuCySGlnAlaAsn--- 39
Db 631 GGAGCTGTATAGTCTCTGATTATGACGATCCAGCTTTCAGCTGATGTGCTCCTA 690
QY 40 ---GlnGlyGlySerSerMetTyrSerGlyProSerSerLeu-----ValTyrThrSer 56
Db 691 GGCATATGAGCGCGCTGTGCTGCTCAGAGAGAGGAGCATCAACACTTACATCCCTCTA 750
QY 57 AlaMetProGlyPheProTyrPro---AlaAlaThrAlaAlaAlaAlaTyrArgGlyAla 75

```

```

Db 751 ATCATTCCTGGGCTTCCCTTACCACTGACGACACACGCGAGCTTTCGAGGAGCC 810
QY 76 HisLeuArgGlyArgGlyArgThrValTyrAsnThrPheArgAlaAlaAlaProProPro 95
Db 811 CATGTGAGGAGGAGAGGAGGAGGAGCAGTGTATGTGTGACAGTCCG---GCGGTACTCCACA 867
QY 96 ProIleProAlaTyrGlyGlyValValTyrGlnGluProValTyrGlyAsnLysLeu 115
Db 868 GCCATCCCGCCCTATCCAGGTGTGTATACCGAGCGGATTTACGGTGTCTGAC---CTC 924
QY 116 GlnGlyGlyTyrAlaAlaTyrArgTyrAlaGlnProThrProAlaThr----- 131
Db 925 TATGTGATATGACAGCCTACAGATATGACAGCCTGCTACTGTACACGCGACACAGCT 984
QY 132 -----AlaAlaAlaTyrSerAspSerTyrGlyArgValTyrAlaAlaAspPro 147
Db 985 GCTGACGCGCGCTGACGCGCTTACAGCAGCGGTTACGGAGGCTGTACACAGTGAACCC 1044
QY 148 TyrHisHisThrLeuAlaProAlaProThrTyrGlyValGlyAlaMetAsnAlaPheAla 167
Db 1045 TAC---CATGCCCTGCGCCCTGCGCCGACGCTATGAGAGTTGGCGCTGTG----- 1089
QY 168 ProLeuThrAspAlaLysThrArgSerHisAlaAspAspValGlyLeuValLeuSerSer 187
Db 1089 -----

QY 188 LeuGlnAlaSerIleTyrGlnGlyTyrAsnArgPheAlaProTyr 203
Db 1090 -----GCGAGTTGTACCGAGGTGCTACAGCCGATTGGCCCCCTAC 1131

```

Search completed: March 16, 2003, 03:31:43
Job time : 1920 secs

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: March 15, 2003, 16:22:47 ; Search time 225 Seconds
(without alignments)
2031.803 Million cell updates/sec

Title: US-09-809-545A-2
Perfect score: 1086
Sequence: 1 MTNKKAVNPYNGWKLNPVV.....VLSSLSQASLYOGYNYREAPY 203

Scoring table:

BLOSUM62			
Xgapop 10.0 , Xgapext 0.5			
Xgapop 6.0 , Xgapext 7.0			
Delop 6.0 , Delext 7.0			

Searched: 2185239 segs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

```

-MODEL=frame+.p2n.model -DEV=xlp
-Q=/cgnr_1/USPTO.spool/US09809545/runat_07032003_153852_11445/app_query.fasta-1.391
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blcsum2 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=DCT -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09809545 @CGN_1_1_396 @runat_07032003_153852_11445 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THRAQS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

Database : N_Geneseq_101002.*

1:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.*
23:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
24:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1086	100.0	1340	AA594693	Rat secreted facto
2	1009.5	93.0	1513	AAH13824	Human cDNA sequenc
3	1009.5	93.0	1800	AAK52245	Human polynucleoti
4	955	87.9	2372	AAA07075	cDNA encoding huma
5	617	56.8	607	AAH09205	Human cDNA clone (
6	606.5	55.8	481	AA593633	DNA encoding novel
7	605	55.7	539	ABK34530	Human cDNA for nov
8	534.5	49.2	567	AA566245	DNA encoding novel
9	534.5	49.2	939	AA161034	Human polynucleoti
10	534.5	49.2	939	AA161035	Human polynucleoti
11	534	49.2	918	AA159249	Human polynucleoti
12	534	49.2	1011	AA159248	Human polynucleoti
13	511	47.1	1252	AA590510	DNA encoding novel
14	467.5	43.0	1506	AA246827	Human RNA binding
15	405.5	37.3	2118	AA593634	DNA encoding novel
16	368.5	33.9	578	AA590506	DNA encoding novel
17	342	31.5	406	AA593632	DNA encoding novel
18	294	27.1	428	ABL68917	Kidney cancer rela
19	258	23.8	1164	AAK53229	Human polynucleoti
20	239	22.0	327	AAK54001	Murine transcripti
21	231.5	21.3	413	AAK04077	Human secreted pro
22	218.5	20.1	407	AA186721	Human polynucleoti
23	198	18.2	316	AAK06335	Human brain expres
24	198	18.2	3189	ABK34602	Human cDNA for nov
25	129	11.9	1104	ABL24461	Drosophila melanog
26	116	10.7	9192	ABL92117	Human Tumour Endoc
27	116	10.7	9483	AA544690	Human full-length
28	116	10.7	11657	ABV23446	Human prostate exp
29	116	10.7	11657	ABV29303	Human prostate exp
30	110.5	10.2	660	ABL07513	Drosophila melanog
31	110.5	10.2	2660	ABL07512	Drosophila melanog
32	110	10.1	109519	AA158069	Microsomopora DNA
33	108.5	10.0	6930	AA158020	Human polynucleoti
34	108.5	10.0	7272	AA159806	Human polynucleoti
35	107	9.9	895	ABN98713	Human polynucleoti
36	103.5	9.5	2834	AAK52977	Human polynucleoti
37	103.5	9.5	4809	AAK51993	Drosophila melanog
38	101.5	9.3	2853	ABL29393	Drosophila melanog
39	101.5	9.3	6906	ABL29392	Drosophila melanog
40	101	9.3	914	AAK34074	Human EST R2810 DN
41	100.5	9.3	2826	ABL05376	Drosophila melanog
42	100	9.2	2564	ABL06127	Drosophila melanog
43	100	9.2	5340	ABL06126	Drosophila melanog
44	100	9.2	12403	ABL07646	Drosophila melanog
45	99.5	9.2	5643	ABK93133	Human prostate spe

ALIGNMENTS

RESULT 1
ID AA594693 standard; cDNA, 1340 BP.
AA594693;
12-MAR-2002 (first entry)
Rat secreted factor DNA clone P0184_D11 #1.
Rat; secreted factor polypeptide; cardiac disease; kidney;
inflammatory disease; congestive heart failure; myocarditis; ashma; ss;
dilated congestive cardiomyopathy; angina pectoris; cardiac arrhythmia;
myocardial infarction; pulmonary hypertension; arteriosclerosis; stroke;
atherosclerosis; cardiac tumour; glomerulonephritis; nephrotic syndrome;
renal infection; hereditary nephritis; polycystic kidney disease;
chronic renal failure; renal vein thrombosis; medullary sponge kidney;
rheumatoid arthritis; osteoarthritis; poriasis; restenosis; PCR primer;

KW graft versus host reaction; Crohn's disease; ulcerative colitis; probe;
KW Alzheimer's disease; gene therapy.
XX Rattus norvegicus.
XX WO200174901-A2.
XX 11-OCT-2001.
XX 23-MAR-2001; 2001WO-US09555.
XX 31-MAR-2000; 2000US-193548P.
XX 14-MAR-2001; 2001US-0809545.
XX (SCIO-) SCIOS INC.
XX Stanton LW, White RT;
XX WPI; 2002-010779/01.
XX P-PsDB; AAU70146.
XX Novel secreted factor polypeptide useful for treating cardiac diseases
XX such as arteriosclerosis, myocardial infarction, inflammatory diseases
XX such as asthma, stroke, and rheumatoid arthritis and renal diseases -
XX
XX Claim 1; Fig 1; 189pp; English.
XX
XX The invention relates to rat secreted factor polypeptides and the
XX polynucleotides encoding them. The sequences are useful for treating
XX cardiac, renal or inflammatory diseases. These include cardiac diseases
XX such as congestive heart failure, myocarditis, dilated congestive
XX cardiomyopathy, angina pectoris, myocardial infarction, cardiac
XX arrhythmia, pulmonary hypertension, arteriosclerosis, atherosclerosis and
XX cardiac tumours, renal diseases such as glomerulonephritis, nephrotic
XX syndrome, renal infarction, hereditary nephritis, polycystic kidney
XX disease, chronic renal failure, renal vein thrombosis and medullary
XX sponge kidney and inflammatory diseases such as asthma, rheumatoid
XX arthritis, osteoarthritis, stroke, psoriasis, restenosis, graft versus
XX host reaction, Crohn's disease, ulcerative colitis and Alzheimer's
XX disease. Sequences AAS94693-AAS94745 represent cDNA clones, which encode
XX the secreted factor polypeptides of the invention, and oligonucleotide
XX probes and PCR primers.
XX
XX Sequence 1340 BP; 366 A; 394 C; 317 G; 263 T; 0 other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 9 436-89 Length: 1340
XX Score: 1086.00 Matches: 203
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 24 Gaps: 0
XX
XX US-09-809-545A-2 (1-203) x AAS94693 (1-1340)
XX
XX 1 MetThrAsnLysLeuAlaValAsnProTyrThrAsnGlyTyrPheLeuAsnProVal 20
XX 535 ATGACGTAATATAAAGCGCGTGAACCCCTACCAATGGCGGAAATTAATCAAGTTGG 54
XX
XX 21 GtAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuGlyGlnAlaGln 40
XX 535 GCGCGGCTTACAGCCCACTTCTATGACGACGCGTGTGTGGCCAGGCAACGAG 654
XX
XX 41 GtGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
XX 655 GAGGAGTCTTCATGATGACGTGCGCCCAAGTTCCTGTGATATATCTTGTCAATGCTGGC 714
XX
XX 61 PheProTyrProAlaAlaThrAlaAlaAlaAlaTyrArgGlyValHisLeuArgGlyArg 80
XX 715 TTTCATATCCGCGCGCACTGCTGACGCTGATACGAGGGGCTCACTTGAAGCCGT 774
XX
XX 81 GtAlaGThrValTyrAsnThrPheArgAlaAlaAlaProProProProIleProAlaTyr 100
XX
XX

DB 775 GGTGCAACCGTGTACACACCTTCAGAGCTGGCGCCGCCCAATCCCGGCTTAT 834
XX
XX 101 GtGlyValAlaTyrGlnGlnProValTyrGlyAsnLysLeuLeuGlnGlyTyrAla 120
XX 835 GCGGAGTAGTGTATCAAGCCAGTGTATGGCAATTAATGCTACAGGTTGTTACGCT 894
XX
XX 121 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSerTyrGly 140
XX 895 GCATACCGCTACGCGCCAGCCCACTGCTGCTGCTGCTACAGTACAGTTACGGA 954
XX
XX 141 ArgValTyrAlaAlaAspProTyrHisThrLeuAlaProAlaProThrTyrGlyVal 160
XX 955 CGAGTTATGTCGCGACCCCTACACACACCTTGTCTCAGCCCACTTACGCGCTT 1014
XX
XX 161 GtAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAsp 180
XX 1015 GTGGCATTAATGCTTTTGGCGCTTGACCGATGCCAATAGAGGATGATGAT 1074
XX
XX 181 ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrGlnGlyGlyTyrAsnArgPhe 200
XX 1075 GTGGGCTGCTGTTCTTCTTCAATGACGCTAGTATATACCAAGGGGATACAAACCGTTT 1134
XX
XX 201 AlaProTyr 203
XX 1135 GCTCATAT 1143
XX
XX
XX RESULT 2
XX AAH13824
XX ID AAH13824 standard; cDNA; 1513 BP.
XX
XX AAH13824;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human cDNA sequence SEQ ID NO:10786.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 10786; 2537bp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide and an oligonucleotide comprising a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH1628 and
CC AAH1863 to AAH1874 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH1629 to AAH1632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX
XX
SQ Sequence 1513 BP; 354 A; 418 C; 395 G; 346 T; 0 other;

Alignment Scores:

Pred. No.:	8.6e-82	Length:	1513
Score:	1009.50	Matches:	190
Percent Similarity:	95.57%	Conservative:	4
Best Local Similarity:	93.60%	Mismatches:	8
Query Match:	92.96%	Indels:	1
DB:	22	Gaps:	1

US-09-809-545A-2 (1-203) x AAH1824 (1-1513)

QY 1 MetThrAsnLysLysAlaValAsnProTyrThrAsnGlyTTPDLSLeuAsnProValVal 20
DB 806 ATGCAATAAATAAGACCGCTACACCTTATCAATATGCTGAAATGAAATCCAGTTGTG 865
QY 21 GAlaValAlaTyrSerProAspPheTyrAlaGlyThrValLeuLeuCysGAlaAlaAsnGln 40
DB 866 GGTGACGCTTACAGTCCCAATTTATGACAGCAGCTGCTGTTGTCAGGACCAACAG 925
QY 41 GAluGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMet-ProGly 60
DB 926 GAGGATCTTCATGATACAGTCCGCCAGTTCACTTGTATATCTTCGCAATGCCAGGC 985
QY 61 PheProTyrProAlaAlaThrAlaAlaAlaAlaTyrAlaGlyAlaHisLeuArgGlyArg 80
DB 986 TTCCCGTATCCAGACGACCGCCGCGCCGCTTACGAGGGGCGCACCTCGAGGCGGC 1045
QY 81 GAlaTyrThrValTyrAsnThrPheArgAlaAlaAlaProProProGlyLeuProAlaTyr 100
DB 1046 GGTGACCGCTTACAGTCCCAATTTATGACAGCAGCTGCTGTTGTCAGGACCAACAG 1105
QY 101 GAlaGlyValAlaTyrGlnGluProValTyrGlyAsnLysLeuLeuGlnGlyGlyTyrAla 120
DB 1106 GCGCGTGTGTTTACCGAGATGATTTATGTCAGAC---ATTATGCTGTGTTATCT 1162
QY 121 AAlaTyrAlaGlyTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSerTyrGly 140
DB 1163 GCATATCCGCTACGCGCCACCTTACCTCCACTGCGCTGCTTACAGTACGATACGGA 1222
QY 141 ArgValTyrAlaAlaAspProTyrHisThrLeuAlaProAlaProThrTyrGlyAla 160
DB 1223 CGAGTTATGCTGCGCGACCCCTTACCAACGACTGCTCCAGCCCACTTACGCGCTT 1282
QY 161 GAlaAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAspAsp 180
DB 1283 GGTGCAATGATGCTTTGCACTTGTACTGATGCCAAGACATGAGAGCATGCTGATGAT 1342
QY 181 ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrGlnGlyGlyTyrAsnArgPhe 200
DB 1343 GTGGGCTCGTCTTCTTCAATGACAGCTAGTATATACCGAGGGGATACACCGTCTT 1402
QY 201 AlaProTyr 203
DB 1403 GCTCCATAC 1411
RESULT 3

AAK52245
ID AAK52245 standard; cDNA; 1800 BP.
XX
XX AAK52245;
AC
XX 06-NOV-2001 (first entry)
DT
XX
XX Human polynucleotide SEQ ID NO 790.
DE
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
XX Homo sapiens.
OS
PN NO200157190-A2.
XX
XX
XX 09-AUG-2001.
PD
XX
XX 05-FEB-2001; 2001WO-US04098.
PF
XX
XX 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
PR
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Gao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AD, Yang Y, Wejrtman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
DR P-PSDB; AAM79112.
XX
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX
XX Claim 1; Page 2643-2645; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK51435) and the
CC encoded polypeptides (AAM78323-AAH80302) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX
SQ Sequence 1800 BP; 432 A; 552 C; 481 G; 335 T; 0 other;

Alignment Scores:

Pred. No.:	1.06e-81	Length:	1800
Score:	1009.50	Matches:	190
Percent Similarity:	95.57%	Conservative:	4
Best Local Similarity:	93.60%	Mismatches:	8
Query Match:	92.96%	Indels:	1
DB:	22	Gaps:	1

US-09-809-545A-2 (1-203) x AAK52245 (1-1800)

QY 1 MetThrAsnLysLysAlaValAsnProTyrThrAsnGlyTTPDLSLeuAsnProValVal 20


```

Db 1093 ATGACAAATAAAAGACCTCAACCTTATACAAATGCTGGAAATGAATCAGTTGTG 1152
Qy 21 G1YAlaValTYrSerProAspPheTYrAlaG1YThrValLeuLcYsG1n1a1aAngin 40
Db 1153 GGTGAGCTTACAGTCCCAATTTCTATGAGACGCGTCTGTTGTGCGACGGCCACACAG 1212
Qy 41 GIUGlySerSerMetTYrSerGlyProSerSerLeuValTYrThrSerAlaMetProGly 60
Db 1213 GAGGATCTTCATGTACAGTGGCCCCAGTTCACTGTATATATCTTCTGCAATGCCAGGC 1272
Qy 61 PheProTYrProAlaAlaThraAlaAlaAlaTYrArgGlyAlaHisLeuArgGlyArg 80
Db 1273 TTCCCGTATCCAGACGACCCGCGCGGCTTACCGAGGGCGCACCTGGGAGGCCGC 1332
Qy 81 G1YArgThrValTYrAsnThrPheArgAlaAlaAlaProProProProLleProAlaTYr 100
Db 1333 GGTCCGACCGTATACACACCTTCAGGGCGCGGCGCCCCCGCCGATCCCGGCTTAC 1392
Qy 101 G1YGIYValValTYrG1nG1nLuproValTYrG1YAsnLYLeuLeuG1nG1YTYrAla 120
Db 1393 GCGGCTGTTGTTTACAGAGATGGATTTATGGTCAGAC---ATTATGTTGTTATGCT 1449
Qy 121 AlaTYrArgTYrAlaG1nProThrProAlaThraAlaAlaTYrSerAspSerTYrGly 140
Db 1450 GCATACCGCTACGCCACGCTACCCCTGCACCTGCGCTGCTACAGTACGATTCGGA 1509
Qy 141 ArgValTYrAlaAlaAspProTYrHisH1stHrLeuAlaProAlaProThrTYrGlyVal 160
Db 1510 CAGATTATGCTGCGACCCCTTACACGACGACTTGTCTCCAGCCCCCTTACGGCGTT 1569
Qy 161 G1YAlaMetAsnAlaPheAlaProLeuThrAspAlaAlaTYrArgSerHisAlaAsp 180
Db 1570 GGTGCATGAATGCTTTTGCACCTTGACTGATGCCAAGCTAGAGCCATGCTGATGAT 1629
Qy 181 ValG1YLeuValLeuSerSerLeuG1nAlaSerIleTYrG1nG1YGIYTYrAsnArgPhe 200
Db 1630 GTGGGCTCTGTTCTTTCATGTGACGCTAGATATACCGAGGGGAGATACAAACGTTT 1689
Qy 201 AlaProTYr 203
Db 1690 GCTCCATAC 1698

```

RESULT 4
AAA07075 standard; cDNA; 2372 BP.

```

XX AC AAA07075;
XX DT 03-JUL-2000 (first entry)
XX DE cDNA encoding human ataxin-2 binding protein (A2BP).
XX KW Ataxin-2 binding protein; A2BP; human; RNA-binding; cell polarisation;
XX KW neuronal plasticity; cellular degeneration signal transduction pathway;
XX KW selective RNA transport; spinocerebellar ataxia type-2;
XX KW hyperproliferative disorder; ss.

```

```

OS Homo sapiens.
XX Key Location/Qualifiers
XX CDS 987..1979
XX FT /*tag= a
XX FT /product= "Human A2BP"

```

```

XX WO200012710-A1.
XX PD 09-MAR-2000.
XX PF 01-SEP-1999; 99WO-US20156.
XX PR 01-SEP-1998; 98US-0145391.
XX

```

(CEDA-) CEDARS SINAI MEDICAL CENT.
PI Pulic SM, Shibata H;
XX WPI: 2000-237873/20.
DR P-98DB; AA181462.
XX
PT Nucleic acid encoding an ataxin-2 binding protein useful for
PT inhibiting the expression of active proteins from the SCA2 gene for the
PT treatment of spinocerebellar ataxia type-2 -
XX
PS Claim 6; Page 74-77; 82pp; English.

This sequence represents cDNA encoding human ataxin-2 binding protein (A2BP). Nucleotide sequences encoding human A2BP were originally isolated in an adult brain cDNA library using the yeast two hybrid method. The ligand of A2BP, ataxin-2, is a 40 kD protein of unknown function that is encoded by the SCA2 gene located on chromosome 12. SCA2 has been linked to the autosomal dominant neurodegenerative disorder spinocerebellar ataxia type-2. Individuals afflicted with the disease exhibit CAG triplet expansion in the SCA2 gene, resulting in ataxin-2 containing a polyglutamine stretch of about 35-39 residues, whereas that of normal individuals contains approximately 22 contiguous glutamine residues. A2BP and ataxin-2 are components of a cellular degeneration signal transduction pathway. The pathogenic expanded form of ataxin-2 has a higher affinity for A2BP relative to normal ataxin-2; the presence of the expanded form is likely to promote degeneration. A2BP and ataxin have also been found to have a role in gene regulation. The binding of A2BP to ataxin-2 plays an important role in controlling gene expression via the targeting of transport of specific RNAs, selective RNA transport being mediated via the RNA binding domains of A2BP. A2BP is expressed very early in embryonic development. Both ataxin-2 and A2BP are able to bind CC RNA, and are essential components of the RNA localisation network that establishes cellular polarity in embryogenesis. In highly differentiated, polarised cells such as neurons, A2BP and ataxin-2 have a similar function and are required for neuronal plasticity. A2BP nucleic acids may be used for the recombinant production of A2BP proteins or fragments thereof according to standard methodologies. For example, an A2BP protein CC with an ataxin-2 or RNA binding capability but no signal transduction CC function can be used as a dominant negative inhibitor of the cellular degeneration signal transduction pathway. A2BP proteins with a signal transduction function can be used to treat hyperproliferative disorders (e.g., cancer) via stimulation of the cellular degeneration pathway.

Sequence 2372 BP; 548 A; 726 C; 646 G; 452 T; 0 other;

Alignment Scores:
Pred. No.: 1,21e-76 Length: 2372
Score: 955.00 Matches: 189
Percent Similarity: 86.94% Conservative: 4
Best Local Similarity: 85.14% Mismatches: 9
Query Match: 87.94% Indels: 21
DB: 21 Gaps: 2

US-09-809-545A-2 (1-203) x AAA07075 (1-2372)

```

Qy 1 MetTYrAsnLYLeuAlaValAsnProTYrThrAsnG1YTYrLYLeuAsnProValVal 20
Db 1572 ATGACAAATAAAAGACCTCAACCTTATACAAATGCTGAAATTCATGATTCAGTTG 1631
Qy 21 G1YAlaValTYrSerProAspPheTYrAlaG1YThrValLeuLcYsG1n1a1aAngin 40
Db 1633 GGTGAGCTTACAGTCCCAATTTCTATGAGACGCGTCTGTTGTGCGACGCAACAG 1691
Qy 41 GIUGlySerSerMetTYrSerGlyProSerSerLeuValTYrThrSerAlaMetProGly 60
Db 1692 GAGGATCTTCATGTACAGTGGCCCCAGTTCACTGTATATATCTTCTGCAATGCCAGGC 1751
Qy 61 PheProTYrProAlaAlaThraAlaAlaAlaTYrArgGlyAlaHisLeuArgGlyArg 80
Db 1752 TTCCCGTATCCAGACGACCCGCGCGGCTTACCGAGGGCGCACCTGGAGGCCGC 1811
Qy 81 G1YArgThrValTYrAsnThrPheArgAlaAlaAlaProProProProLleProAlaTYr 100

```



```
Db 1812 GGTGGACCGGTATACACCTTCAGGGCCGCGGCCGCCGATCCCGGCTTAC 1871
Qy 101 G1G1G1ValValTyr-G1ng1uProValTyrG1Yasn1Ys1euleuG1ng1Y1YrA1 120
Db 1872 GCGGCTGTGTTTACCCAGGATGATTTATGTCACAC--ATTATGTTGTTATGC 1928
Qy 120 aAlTyrArgTyrAlaG1nProThrProAlaThrAlaAlaTyrSerAspSer----- 138
Db 1929 TGCATACCGGTACGCGCCAGCCCTACCCCTGCGCATCGCTGACGTACAG-AAATC 1987
Qy 139 -----TyrG1YArVa 142
Db 1988 AGTTCGCTTCCTGTCACAGATGAATTTCTTGTACACCTCTGACAGTTCCGACGAGT 2047
Qy 142 1TyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrG1YValG1YAl 162
Db 2048 TTTATGCTGCCAACCCTTACACACGACGCTGCTCAGCCGCCCTTACGGCGTTGGTGC 2107
Qy 162 aMetAsnAlaPheAlaProLeuThrAspAla1YsThrArgSerHisAlaAspAspValG1 182
Db 2108 CATGAATGCTTTTGACCTTGTGACTGATGCCAAGACTAGAGCCATGCTGATGATGTCGG 2167
Qy 182 YLeuVal1euSerSerLeuG1nAlaSer11eTyrG1ng1Y1YrAsnArgPheAlaPr 202
Db 2168 TCTGCTCTTCTTTCATGACAGAGGTATATACCGAGGGGATACCAACCGTTTGTCTCC 2227
Qy 202 cTyr 203
Db 2228 ATAC 2231
```

RESULT 5

AAH09205/c
ID AAH09205 standard; cDNA; 607 BP.

AC AAH09205;

DT 26-JUN-2001 (first entry)

DE Human cDNA clone (3'-primer) SEQ ID NO:6040.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

PI Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

PS Claim 3; SEQ ID 6040; 2537bp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB93893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 607 BP; 140 A; 152 C; 179 G; 129 T; 7 other:

Alignment Scores:
Pred. No.: 6, 34e-47 Length: 607
Score: 617.00 Matches: 136
Percent Similarity: 83.04% Conservative: 6
Best Local Similarity: 79.53% Mismatches: 21
Query Match: 56.81% Indels: 8
Gaps: 2

US-09-809-545A-2 (1-203) x AAH09205 (1-607)

```
Qy 40 G1ng1uG1Ys1eSerMetYrSerG1Y---ProSerSerLeuValTyrThrSerAla-Me 58
Db 600 CAGAGGAGNTTTTTCATGAGCAGAGCCGCCAGTTCATCTGGAANATTTTGTCAANT 541
Qy 58 tProG1YPhePro-TyrProAlaAlaThrAlaAlaAlaTyrArgG1YAlaHisLeuA 78
Db 540 GCCAGGCTTNCGGTATTCAGAGAGCCACCGCGCGGCGCTTACCGAGGGCGCAGCTTGC 481
Qy 78 tG1YArVaG-G1YArGThrValTyrAsnThrPheArgAlaAlaAla-PropProProPoi1 97
Db 480 GAGNCCGGGGGTCCGACCGTGTACACACCTTCAGGGCGCGCGCCGCCCGAT 421
Qy 97 ePro-AlaTyrG1Y-G1YValValTyrG1ng1uProValTyrG1Yasn1Ys1eLeuG1n 116
Db 420 CCGGGCCCTACGGCGGGGTGTTTACCGAGGATTTATGAGTGCAGAC--ATTAT 364
Qy 117 G1G1Y1YrAlaAlaTyrArgTyrAlaG1nProThrProAlaThrAlaAlaAlaTyrSer 136
Db 363 GGTGTTATGCTGCATACCGGTACGCCAGCTTACCCCTGCACTGCCGCTGCTACAGT 304
Qy 137 AspSerTyrG1YArVa1YrAlaAlaAspProTyrHis1ThrLeuAlaProAlaPro 156
Db 303 GACAGTTACGAGCAGATTATGTCGCGACCCCTACACACGACGACCTGCTCCAGCCCC 244
Qy 157 ThrTyrG1YValG1YAlaMetAsnAlaPheAlaProLeuThrAspAla1YsThrArgSer 176
Db 243 ACCTACGCGGTGTGTCATGAATGCTTTTGACCTTGTGATGTCGACAGCAGAGGAGC 184
Qy 177 His1AlaAspAspValG1YLeuVal1euSerSerLeuG1nAlaSer11eTyrG1ng1Y1Y 196
Db 183 CATGCTATGATGTGGGTCTGCTCTTCTTCATTCGACGCTAGTATATACGAGGGGA 124
Qy 197 TyrAsnArgPheAlaProTyr 203
Db 123 TACAACGTTTGTCTCATAC 103
```

RESULT 6

AAH93633
ID AAH93633 standard; cDNA; 481 BP.

XX

AC AAS93633;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #29437.
XX
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
PD 11-OCT-2001.
PP 30-MAR-2001; 2001WO-US08631.
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
PI WPI: 2001-639362/73.
DR P-P9DB; ABS29446.
XX
PS Claim 1; SEQ ID NO 29437, 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (I) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostic, forensic, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX
CC
CC Sequence 481 BP; 100 A; 145 C; 123 G; 113 T; 0 other;

Alignment Scores:

Pred. No.:	Length:
4,64e-46	481
Score: 606.50	Matches: 120
Percent Similarity: 90.51%	Conservative: 4
Best Local Similarity: 87.59%	Mismatches: 10
Query Match: 55.85%	Indels: 3
DB: 23	Gaps: 1

US-09-809-545A-2 (1-203) x AAS93633 (1-481)

69 AATAAATATYRAGGJYALAHLEUARGJLYRGJYAKTNRVALTYRASNTHRPHE 88
Db 1 GCCGGCGCCTTAACGAGGGGCGCACTTCGAGGCCGCCGTGCACCTGTACAACACTTC 60
89 ATGATGAATAAATPOTPTPTPTTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 108

Db 61 AGGGCGGGGGCCCCCCCCCGGCTACCGGGCTACGGCGGTGTGTTTACCAGGATGGA 120
Qy 109 VAlTYrGlyAsnLysLeuLeuGlnGlyGlyTyrAlaAlaTyrArgTyrAlaGlnProThr 128
Db 121 TTTTATGGTGACAGAC--ATTATATGTGGTTATVGTGTCATACCGCTACGCCCGCTAC 177
Qy 129 ProAlaThrAlaAlaAlaTyrSerAspSerTyrGlyArgValTyrAlaAla--AspProT 148
Db 178 CTTGCACCTGCGCGCTGCTCTACGTACGTACACTTACGACAGATTATGACATTTTCCCCCT 237
Qy 148 YrHAlaHThrLeuAlaProAlaProThrTyrGlyValGlyAlaMetAsnAlaPheAlaP 168
Db 238 GTCACACGACTCAGCTCTCCAGCCCCACCTACCGGCGTGTGTCCTCATGATGCTTTTGAC 297
Qy 168 IoleuThrAspAlaLysThrArgSerHisAlaAspAspValGlyLeuValLeuSerSerL 188
Db 298 CTTTACCTGATGTCCAAGACTAGAGCCACTGCTGATGATGTGGGTCTCGTTCTTTCTTCA 357
Qy 188 euGlnAlaSerLysTyrGlnGlyGlyTyrAsnArgPheAlaProTyr 203
Db 358 TGCAGGCTAGTATATATCCAGGGGATACACCGTTTGTCTCCATAC 404
RESULT 7
ID ABK34530/c
ID ABK34530 standard; cDNA; 539 BP.
XX ABK34530;
AC
XX
XX
DT 08-MAY-2002 (first entry)
XX
DE Human CDNA for novel secreted protein, SEQ ID 299.
XX
XX Human; ss; gene; secreted protein; immune deficiency; viral infection;
KW bacterial infection; fungal infection; autoimmune disorder; burn;
KW rheumatoid arthritis; multiple sclerosis; autoimmune thyroiditis;
KW diabetes; allergy; asthma; neurodegenerative disease; Crohn's disease;
KW Alzheimer's disease; Parkinson's disease; liver fibrosis; tumour;
KW coagulation disorder; haemophilia; inflammatory disorder; ulcer;
KW tissue regeneration; wound healing; haematopoiesis; myeloid deficiency;
KW lymphoid cell deficiency.
XX
XX Homo sapiens.
OS
XX
XX WO200177290-A2.
PN
XX
XX 18-OCT-2001.
PD
XX
XX 29-MAR-2001; 2001MO-US10295.
PF
XX
XX 06-APR-2000; 2000US-194941P.
PR
XX
XX (GEMY) GENETICS INST INC.
PA
XX
XX Mong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
PI Gulukota K, Graham JR;
DR
XX WPI; 2002-179323/23.
XX
PT Six hundred and twenty five polynucleotides derived from a variety of
PT human tissue sources which encode secreted proteins, useful for
PT treating immune deficiencies and disorders such as autoimmune disorders
PT
PS
XX
XX
XX Claim 1, Page 153, 339pp; English.
XX
XX The invention relates to 625 polynucleotides which have been derived from
CC a variety of human tissue sources and which encode novel secreted
CC proteins, their complements and sequences that hybridise to them.
CC Also included are a vector comprising the polynucleotide, a host cell
CC transformed with the vector, the proteins encoded by the
CC polynucleotides, antibodies that bind to the proteins and identification
CC of modulators of the proteins or the expression of the polynucleotide.


```

Db 402 TTCCCTACCCACCCGACAGCCGCTTCCGCGGCGGACATCTTGGGGCCGG 343
Qy 81 GYATGTTThValTYrAsnThrPheArgAlaIaIaProProProIleProAlaTYr 100
Db 342 GGGCGGGCGGTATATATATATATATGCGGCTGGCCACCCCGCATCCCGACTTAC 283
Qy 101 GYGYValValTYrGInGluProValTYrGlyAsnLYsLeuLeuGInGlyTYrAla 120
Db 282 GGAGCGGTGCTGTATCAGATGATTTATGTGTGAG---ATTATGAGGCTACGCA 226
Qy 121 AATYrArgTYrAlaGInProThrProAlaThrAlaIaIaTYrSerAspSerTYrGly 140
Db 225 GCGTACAGATAGCTCTACGCCCTCT---GGAGCGGCGGCGCTTACGCGACGTTACGCGC 169
Qy 141 ArgValTYr---AlaIaAspProTYrHissthrLeuAlaProAlaProThrTYrGly 159
Db 168 AAGTCTACGCGAGCTGCCACCGGTACATCACACCATCGGCGCGGCGACTTACAGC 109
Qy 160 ValGlyAlaMetAsnAlaPheAlaPro----- 168
Db 108 ATTGGAACCATGTGAACCTTCCACCGCTTCTTCGAGACCATGAAGGCAAAACAA 49
Qy 169 -----LeuThrAspAlaLYsThrArgSerHis 177
Db 48 AAACAAAAAAATATCACAACAAAAACAAAAACAAACAC 10

RESULT 9
AAI61034/c
ID AAI61034 standard; cDNA; 939 BP.
XX
AC AAI61034;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 5023.
XX
KW Human; nocotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang J, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR MPI; 2001-44253/47.
XX P-PsDB; AAM41878.
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries
XX

```

```

PS Claim 1; SEQ ID NO 5023; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM3642-AAM4221) with nocotropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, cancer diagnosis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 939 BP; 148 A; 248 C; 297 G; 246 T; 0 other:
XX
Alignment Scores:
Pred. No.: 3,22e-39 Length: 939
Score: 534.50 Matches: 109
Percent Similarity: 60.62% Conservative: 8
Best Local Similarity: 56.48% Mismatches: 31
Query Match: 49.22% Indels: 45
DB: Gaps: 22
XX
US-09-809-545A-2 (1-203) x AAI61034 (1-939)
Qy 1 MetThrAsnLYsLYsAlaValAsnProTYrThrAsnGlyTYrLYsLeuAsnProValVal 20
Db 501 ATGACCAACAAGAAAGACGGGAGACCCCTTACACCAACGCGCTGAACTAAATCCATGTGTC 442
Qy 21 GYAlaValAlTYrSerProAspPheTYrAlaGlyThrValLeuLeuCYsGlnAlaAsnGln 40
Db 441 GCGGAGCTTACGGCCCTGAATTTCTATGACAGTACG----- 406
Qy 41 GYIlySerSerMetTYrSerGlyProSerSerLeuValTYrThrSerAlaMetProGly 60
Db 405 -----GGG 403
Qy 61 PheProTYrProAlaAlaThrAlaAlaIaIaTYrArgLYsAlaHisLeuArgLYsArg 80
Db 402 TTCCCTACCCACCCACCGGACAGCCCGTTCCTACCGGCGGCGACATCTTGGGGCCGG 343
Qy 81 GYATGTTThValTYrAsnThrPheArgAlaIaIaProProProIleProAlaTYr 100
Db 342 GGGCGGGCGGTATATATATATATGCGGCTGGCCACCCCGCATCCCGACTTAC 283
Qy 101 GYGYValValTYrGInGluProValTYrGlyAsnLYsLeuLeuGInGlyTYrAla 120
Db 282 GGAGCGGTGCTGTATCAGATGATTTATGTGTGAG---ATTATGAGGCTACGCA 226
Qy 121 AATYrArgTYrAlaGInProThrProAlaThrAlaIaIaTYrSerAspSerTYrGly 140
Db 225 GCGTACAGATAGCTCTACGCCCTCT---GGAGCGGCGGCGCTTACGCGACGTTACGCGC 169
Qy 141 ArgValTYr---AlaIaAspProTYrHissthrLeuAlaProAlaProThrTYrGly 159
Db 168 AAGTCTACGCGAGCTGCCACCGGTACATCACACCATCGGCGCGGCGGCGACTTACAGC 109
Qy 160 ValGlyAlaMetAsnAlaPheAlaPro----- 168
Db 108 ATTGGAACCATGTGAACCTTCCACCGCTTCTTCGAGACCATGAAGGCAAAACAA 49
Qy 169 -----LeuThrAspAlaLYsThrArgSerHis 177
Db 48 AAACAAAAAAATATCACAACAAAAACAAAAACAAACAC 10

RESULT 10
AAI61035/c

```

ID	AA161035	standard; cDNA; 939 BP.
XX	AA161035;	
AC	AA161035;	
XX		
DT	22-OCT-2001	(first entry)
XX		
DE	Human polynucleotide SEQ ID NO 5024.	
XX		
KW	Human; nocotropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; autoimmune; lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokine; thrombolytic; drug screening; arthritis; inflammation; leukaemia; ss.	
KW	leukaemia; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	MO200153312-A1.	
XX		
PD	26-JUL-2001.	
XX		
PF	26-DEC-2000; 2000MO-US34263.	
XX		
PR	21-JAN-2000; 2000US-0488725.	
PR	25-APR-2000; 2000US-0552317.	
PR	09-JUL-2000; 2000US-0598042.	
PR	19-JUL-2000; 2000US-0620312.	
PR	03-AUG-2000; 2000US-0653450.	
PR	14-SEP-2000; 2000US-0662191.	
PR	19-OCT-2000; 2000US-0693036.	
PR	29-NOV-2000; 2000US-0727344.	
XX		
PA	(HYSE-) HYSEQ INC.	
XX		
P1	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;	
P1	Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;	
P1	Zhao QA, Zhou P, Goodrich R, Drmanac RT;	
XX		
DR	WPI; 2001-442253/47.	
XX		
DR	P-PSDB; AAM41879.	
XX		
PT	Novel nucleic acids and polypeptides, useful for treating disorders	
PT	such as central nervous system injuries -	
XX		
PS	Claim 1; SEQ ID NO 5024; 10078bp; English.	
XX		
CC	The invention relates to human nucleic acids (AA157798-AA161369) and	
CC	the encoded polypeptides (AAM38642-AAM42213) with nocotropic,	
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful	
CC	in gene therapy. A composition containing a polypeptide or polynucleotide	
CC	of the invention may be used to treat diseases of the peripheral nervous	
CC	system, such as peripheral nervous injuries, peripheral neuropathy and	
CC	localised neuropathies and central nervous system diseases, such as	
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic	
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the	
CC	utilisation of the activities such as: Immune system suppression,	
CC	Activin/inhibin activity, chemotactic/chemokine activity, haemostatic	
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,	
CC	assays for receptor activity, arthritis and inflammation, leukaemias and	
CC	C.N.S disorders.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification.	
XX		
SO	Sequence 939 BP; 148 A; 248 C; 297 G; 246 T; 0 other;	
XX		
Alignment Scores:		
Pred. No.:	3,22e-39	Length: 939
Score:	534.50	Matches: 109
Percent Similarity:	60.62%	Conservative: 8
Best Local Similarity:	56.48%	Mismatches: 31
Query Match:	49.22%	Indels: 45
DB:	22	Gaps: 5

US-09-809-545A-2 (1-203) x AA161035 (1-939)	
Qy 1	MeThrAnIySylValAlaValAaNPoTYrThrAsnGlyTTrpLysLeuAsnProValAla 20
Db 501	ATGACCAAAAGAAAGACGGGGAACCCCTACACCAACGGCTGGAAGCTTAATCACTAGTGC 442
Qy 21	GIyAlaValIYrSerProAaPheTYrAlaGlyThValLeuLeuCySerIlnAlaenGln 40
Db 441	GGCCAGCTACGAGCGGCTGAATCTATGCAGTGACG----- 406
Qy 41	GIuGlySerSerMetYrSerGlyProSerSerLeuValIYrThrSerAlaMetProGly 60
Db 405	-----GGG 403
Qy 61	PheProTYrProAlaAlaThrAlaAlaAlaIYrArgGlyAlaHisLeuArgGlyArg 80
Db 402	TTCCCTACCCACCAACGCGACAGCCGTTGCTTACCGGGGCGACATCTTCGGGGCGG 343
Qy 81	GIyArgThrValIYrAsnThrPheAlaGAlaAlaAlaProProProProIleProAlaIYr 100
Db 342	GGCGGGGCGGTATTAATACATTTCCGGGCGCGCACCCCCACCCCCATCCCGACTTAC 283
Qy 101	GIyGlyValValIYrGlnGlnProValIYrGlyAsnLysLeuLeuGlnGlyGlyTYrAla 120
Db 282	GGAGCGGCTCGTGTACGAGTGAATTTTATGTGCTGAG--ATTATGAGGCTTACGCA 226
Qy 121	AlaTYrArgTYrAlaGlnProThrProAlaAlaAlaAlaAlaTYrSerAspSerTYrGly 140
Db 225	GCTTACAGATATCGCTCAGCCCGCT---GCAAGGGGGGAGCGCTACAGCGACAGTTACGCC 169
Qy 141	AlYValIYr---AlaAlaAspProTYrHisIleThrLeuAlaProAlaProThrTYrGly 159
Db 168	AGAGCTACAGCAGCTGCGCGACCGGTACCATCACACCATCGGGCGCGCGACTTACAGC 109
Qy 160	ValGlyAlaMetAsnAlaPheAlaPro----- 168
Db 108	ATTGGAACCATGTGAACCTTCCACCGTTTCTTTCGAGCATGAGAGGCAAAAACAA 49
Qy 169	-----LeuThrAspAlaLysThrArgSerHis 177
Db 48	AAAAAATCAACAAAACAAAAAACAATAAACAC 10
RESULT 11	
AA159249	
ID	AA159249 standard; CDNA; 918 BP.
AC	AA159249;
XX	
DT	22-OCT-2001 (first entry)
XX	
DE	Human polynucleotide SEQ ID NO 1452.
XX	
XX	Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer
XX	peripheral nervous system; neuropathy; central nervous system; CNS;
XX	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX	chemokine; thrombolytic; drug screening; arthritis; inflammation;
XX	leukaemia; BS.
OS	Homo sapiens.
PN	W020015312-A1.
XX	
PD	26-JUL-2001.
XX	
PF	26-DEC-2000; 2000WO-US34263.
XX	
PR	21-JAN-2000; 2000US-0488725.
XX	
PR	25-APR-2000; 2000US-0552317.
XX	
PR	09-JUL-2000; 2000US-0598042.
XX	
PR	19-JUL-2000; 2000US-0620312.
XX	
PR	03-AUG-2000; 2000US-0653450.
XX	
PR	14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Aundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao Qa, Zhou P, Goodrich R, Dermanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR P-PSDB: AAM40093.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1; SEQ ID NO 1452; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AA42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 918 BP; 235 A; 300 C; 235 G; 148 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 3,476-39 Length: 918
 Score: 534.00 Matches: 105
 Percent Similarity: 67.68% Conservative: 6
 Best Local Similarity: 64.02% Mismatches: 23
 Query Match: 49.17% Indels: 30
 DB: 22 Gaps: 4
 US-09-809-545a-2 (1-203) x AA159249 (1-918)
 QY 1 MetThrAsnLysValAlaValAlaProTyrThrAsnGlyTrrpLysLeuAnProValVal 20
 DB 438 ATGACCACCAAGAGAGCGGGAAACCCCTACACCAACGCGTGAAGCTAAATTCAGTGGTC 497
 QY 21 G1YAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuGlyGlnAlaGln 40
 DB 498 GGGCGAGCTACCGCGCTGAATTCATGCACTACG----- 533
 QY 41 G1nglySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
 DB 534 -----GGG 536
 QY 61 pheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyValHisLeuArgGlyArg 80
 DB 537 TTCCTCTACCCACCAACCGGACCGCTGGCTACCGGGGGCCACATCTTCGGGGCGG 596
 QY 81 G1ArgThrValTyrAsnThrPheArgAlaAlaAlaProProProProTrrpAlaTyr 100
 DB 597 GGGCGGGCGGTGATTAATCATTTCCGGCTGCCACACCCCACTCCGACTTAC 656
 QY 101 G1GlyValValTyrGlnGlnProValTyrGlyAsnLysLeuLeuGlnGlyTyrAla 120
 DB 657 GGAAGCGGTGCTGATCAGATGATTTTATGTCGTAG--ATTATGAGAGGCTACGCA 713
 QY 121 AlATyrArgTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSerTyrGly 140
 DB 714 GCCTACAGATACCTCAGCCCGCT--GCAGCGCGGCGACCTTACAGCGACGTTACGGC 770

QY 141 ArgValTyr---AlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGly 159
 DB 771 AGAGTCTACGACGCTGCCGACCCGTCATCATCACATCGGGCCGTCGACCTACGACG 830
 QY 160 ValGlyAlaMet 163
 DB 831 ATTGAACCAATG 842
 RESULT 12
 AA159248
 ID AA159248 standard; cDNA; 1011 BP.
 XX
 AC AA159248;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 1451.
 XX
 KM Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KM peripheral nervous system; neuropathy; central nervous system; CNS;
 KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KM chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KM leukaemia; se.
 XX
 OS Homo sapiens.
 XX
 PN MO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Tang YT, Liu C, Aundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang Z, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao Qa, Zhou P, Goodrich R, Dermanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR P-PSDB: AAM40092.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1; SEQ ID NO 1451; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AA42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX SQ Sequence 1011 BP; 260 A; 314 C; 267 G; 170 T; 0 other;
Alignment Scores:
Pred. No.: 3.9e-39 Length: 1011
Score: 534.00 Matches: 105
Percent Similarity: 67.68% Conservative: 6
Best Local Similarity: 64.02% Mismatches: 23
Query Match: 49.17% Indels: 30
DB: 22 Gaps: 4
US-09-809-545A-2 (1-203) x AAI59248 (1-1011)
Qy 1 MetThraenLysAlaValaAsnProTyThrAenglyTrrpLysLeuAsnProVala 20
Db 531 ATGACCAATAGAGAGAGCGGGAGACCCCTACACCAACGGCTGGAACCTAATCATCTGGTTC 590
Qy 21 GLyAlaValaIYrSerProAspPheTyraGlyThrValLeuLeuCyGlnAlaEngln 40
Db 591 GCGCGAGCTCAGCGGCTGAAATTCATATCAGTACG----- 626
Qy 41 GLuGlySerSerMetTySerGlyProSerSerLeuValTYrThrSeraLameProGly 60
Db 627 -----GSG 629
Qy 61 PheProTyProAlaAlaThrAlaAlaAlaAlaTYrArgGlyAlaHisLeuArgGlyArg 80
Db 630 TTCCCTTACCCACACCGGCGACACCGCTTGCCTACCGGGGGGACATCTTGGGGCGG 689
Qy 81 GLyArgThrValTYrAsnThrPheArgAlaAlaAlaProProProProIleProAlaTYr 100
Db 690 GGCGCGGCGGTATTAATACATTTGGGCTGGCGACCCCGCCCATCCCACTTAC 749
Qy 101 GLyGlyValaValaTYrGlnGluProValaTYrGlyAsnLysLeuGlnGlyGlyTYrAla 120
Db 750 GGAGCGGTGTATCAGATGATTTATGTGCTGAG--ATTATGAGGCTTACGCA 806
Qy 121 AlaTYrArgTYrAlaGlnProThrProAlaThrAlaAlaAlaTYrSerAspSerTYrGly 140
Db 807 GCTTACATATGAGTACGCTCAGCCGCT---GCAGGGGGCGGACGCTTACAGCAGTACGGC 863
Qy 141 ArgValaIYr---AlaAlaAspProTyRHisThrLeuAlaProAlaProThrTYrGly 159
Db 864 AGAGCTATGAGCGTGGCGACCCGTACCATCACCATCGGGCGGCGACCTTACAGC 923
Qy 160 ValGlyAlaMet 163
Db 924 ATTGGAACCATG 935
RESULT 13
ID AAS90510 standard; CDNA: 1252 BP.
XX AAS90510;
AC AAS90510;
XX 13-FEB-2002 (first entry)
DT 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #26314.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
OS Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX

PA (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
DR P-PSDB; ABG26323.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX Claim 1; SEQ ID No 26314; 103pp; English.
PS
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polypeptide chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1252 BP; 330 A; 338 C; 324 G; 260 T; 0 other;
Alignment Scores:
Pred. No.: 5.98e-37 Length: 1252
Score: 511.00 Matches: 119
Percent Similarity: 61.47% Conservative: 15
Best Local Similarity: 54.59% Mismatches: 40
Query Match: 47.05% Indels: 46
DB: 23 Gaps: 7
US-09-809-545A-2 (1-203) x AAS90510 (1-1252)
Qy 1 MetThraenLysAlaValaAsnProTyThrAenglyTrrpLysLeuAsnProVala 20
Db 563 ATGACCAATAGAGAGAGTGTACACACATATGCTTGAATTAAGCCAGTACTT 622
Qy 21 GLyAlaValaIYrSerProAspPheTyraGlyThrValLeuLeuCyGln----- 37
Db 623 GGAGCTGTATGTGCTCGGAGCTTATATGACATCCAG-CTTATCAGCAATGTCGCC 681
Qy 38 -----AlaEnglnGlnGlySerSerMetTySerGlyProSerSerLeuValTYrThr 55
Db 682 TAGGCATGATGACAGCGCCCTATCAGAGAGGGGGTATCAACCTTACATCTT 741
Qy 56 SeraLameProGlyPheProTyPro---AlaAlaThrAlaAlaAlaTYrArgGly 74
Db 742 TA-ATCATCTCTGCTTCCCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 800
Qy 75 AlaHisLeuArgGlyArgGlyArgTYrValaTYrAsnThrPheArgAlaAlaAlaProPro 94
Db 801 GCGCATTTGAGGGGCGAGGCGGACGATATATGTCAGTCCG---GCGGTACTTCA 857
Qy 95 ProProIleProAlaTYrGlyGlyValaValaTYrGlnGluProValaTYrGlyAsnLysLe 114
Db 858 ACAGCATCCCGGCTTACCGAGGTGTGTTTACAGGACGAGATTTTACGGTCTGAC-- 915
Qy 114 uLeuGlnGlyGlyTYrAlaAlaTYrArgTYrAlaGlnProThrProAlaThr----- 131


```

Db 916 -CTTATGCTGATATGACGCTACGATATGACAGCCTGCTACTGCAACCGCAGCCAC 974
Oy 112 -----AlAlAlAlatYSerApsSerYrYglYArYValYrYAlAlAla 146
Db 975 CCCTGCTGACGCCGCTGACGCCGCTTACGATGACGCTTATGCGAGGCTTACAGCCCA 1034
Oy 146 pProTYrHsHsHrHLeuAlaProAlaProThrTYrYglYValYAlaMeCAsnaLaPh 166
Db 1035 CCCCTAC---CATGCCCTTGCCCTCCGCTGACGATGAGGTGGCGCTGTG----- 1083
Oy 166 eAlaProLeuThrAraPalalYThrYrYrSerHsAlaAraPaValYglYLeuValLeu 186
Db 1083 ----- 1083
Oy 166 rSerLeuGlnAlaSerIleTYrGlnYglYTYrAraArpHeAlaProTYr 203
Db 1084 -----GGAGTTTATACCGAGGTGGCTTACAGCCCATTTGGCCCTTAC 1125

RESULT 14
AAZ46827
ID AAZ46827 standard; cDNA; 1506 BP.
XX
XX AAZ46827;
AC
XX
XX 11-APR-2000 (first entry)
DT
XX
XX Human RNA binding protein (RNABP)-2 encoding cDNA (clone 1250374).
DE
XX
XX RNA binding protein; RNABP; cancer; immune disorder; AIDS; human;
XX
XX developmental disorder; acquired immunodeficiency syndrome; RNABP-2;
XX
XX inflammation; allergy; diabetes mellitus; seizure disorder; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX CDS 201..1301
XX FT /tag= a
XX FT /product= "RNABP-2"
XX
XX
XX US6020164-A.
XX
XX 01-FEB-2000.
XX
XX 21-OCT-1998; 98US-0176657.
XX
XX 21-OCT-1998; 98US-0176657.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Bandman O, Corley NC, Lu DAM, Baughn MR, Tang YT, Guegler KJ;
XX
XX MPI: 2000-146885/13.
XX
XX P-PSDB; AAY56850.
XX
XX Isolated and purified polynucleotide for modulating the expression of
XX
XX human RNA binding proteins which play a role in cancer; immune
XX
XX disorders and developmental disorders -
XX
XX Examples; Columns 51-52; 39pp; English.
XX
XX The present invention provides human RNA binding proteins (RNABP) and
XX
XX polynucleotides encoding the proteins. The polypeptides are useful for
XX
XX modulating the expression of human RNA binding proteins (RNABP) which
XX
XX play a role in cancer; immune disorders and developmental disorders.
XX
XX Disorders associated with a decrease of RNABP include: cancers such as
XX
XX adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma,
XX
XX teratocarcinoma, and, in particular, cancers of the adrenal gland,
XX
XX bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia,
XX
XX gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary,
XX
XX pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen,
XX
XX testis, thymus, thyroid, and uterus; immune disorders such as acquired
XX
XX immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory

```

```

CC distress syndrome, allergies, ankylosing spondylitis, amyloidosis,
CC anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune
CC thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's
CC disease, atopic dermatitis, dermatomyositis, diabetes mellitus,
CC emphysema, episodic lymphopenia with lymphocytotoxicity, erythroblastosis
CC fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis,
CC Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis,
CC hyperostosis, irritable bowel syndrome, multiple sclerosis,
CC myasthenia gravis, myocardial or pericardial inflammation,
CC osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis,
CC Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjogren's
CC syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic
CC sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner
CC syndrome, complications of cancer, hemodialysis, and extracorporeal
CC circulation, viral, bacterial, fungal, parasitic, protozoal, and
CC hematonic infections, trauma, X-linked agammaglobulinemia of Bruton,
CC common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic
CC hypoplasia), thymic dysplasia, isolated immunoglobulin (Ig)-A deficiency,
CC severe combined immunodeficiency disease (SCID), immunodeficiency with
CC thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi
CC syndrome, chronic granulomatous diseases, hereditary angioneurotic edema,
CC and immunodeficiency associated with Cushing's disease; and developmental
CC disorders such as renal tubular acidosis, anemia, Cushing's syndrome,
CC achondroplastic dwarfism, Duchenne and Becker muscular dystrophy,
CC epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia,
CC genitourinary abnormalities, and mental retardation), Smith-Magenis
CC syndrome, myelodysplastic syndrome, hereditary mucopolysaccharidic dysplasia,
CC hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-
CC Tooth disease and neurofibromatosis, hypochromidism, hydrocephalus,
CC seizure disorders such as Sydenham's chorea and cerebral palsy, spina
CC bifida, anencephaly, craniofacial dysmorphism, congenital glaucoma, cataract,
CC sensorineural hearing loss, and any disorder associated with cell
CC growth and differentiation, embryogenesis, and morphogenesis involving
CC any tissue, organ, or system of a subject, e.g., the brain, adrenal
CC gland, kidney, skeletal or reproductive system. The present sequence
CC represents a cDNA encoding the RNABP-2 polypeptide.
XX
XX SQ Sequence 1506 BP; 452 A; 362 C; 377 G; 315 T; 0 other:
XX
XX Alignment Scores:
XX Pred. No.: 6.24e-33 Length: 1506
XX Score: 467.50 Matches: 107
XX Percent Similarity: 56.48% Conserved: 15
XX Best Local Similarity: 49.54% Mismatches: 39
XX Query Match: 43.05% Indels: 56
XX DB: Gaps: 8
XX
XX US-09-809-545A-2 (1-203) x AAZ46827 (1-1506)
Oy 1 MetThrAenLYeAlaValasnProTYrThraenGlyTTPYLeuAasnProValVal 20
Db 768 ATGACCAATPAAAGAGATGGTTCACACCATATGCAATGTGGAAATTTAAGCCAGTAGTT 827
Oy 21 GYAlaValTYrSerProAspHeTYrAlaGlyThrValLeuLeuCYeGlnAlaLaa 39
Db 828 GAGCGTATATGCTCGGAGTTATATGACGATTCAGCTTCAAGCAGATGTGCCA 887
Oy 40 ---GlnGlnGlySerSerMetTYrSerGlyProSerSerLeu-----ValTYrThrSer 56
Db 888 GGCATATGATGACAGAGTGCCTCATCAGAGAGGGGGATCAACATTAATTCCTTAA 947
Oy 57 AlameProGlyPheProTYrPro---AlAlAlatrAlAlAlAlAlAlAlAlAlAlAl 75
Db 948 ATCATTCCTGGCTTCCTTACCTTACCTGACGACACACGCGCGCTTTCAGAGAGCC 1007
Oy 76 HsLeuArGlyArGlyArGrHrValTYrAenThrPheArGAlAlAlAlAlAlAlAlAlAl 95
Db 1008 CATTTGAGGGGAGAGGGGAGAGATATGATGATGATGATGATGATGATGATGATGAT 1064
Oy 96 ProLeuProAlaTYrGlyGlyValValTYrGlnGlnProValTYrGlyAsnHsLeuLeu 115
Db 1065 GCATTCCTGGCTTCCTTACCTTACCTGACGACACACGCGCGCTTTCAGAGAGCC 1081

```


QY 116 GlnGlyGlyTyrAlaAlaTyrArgTyrAlaGlnProThrProAlaThr----- 131
Db 1082 AGGGGTGATGAGCGCTTACAGATGACAGCGCTGTACTGCAACCGCAGCCCGCT 1141
QY 132 -----AlaAlaAlaTyrSerAspSerTyrGlyArgValTyrAlaAlaAspPro 147
Db 1142 GCTCAGCGCTGACCGCTTACAGTACGCTTATGCGAGGCTGTACAGCCGACCC 1201
QY 148 TyrHisThrLeuAlaProAlaProThrTyrGlyValGlyAlaMetAsnAlaPheAla 167
Db 1202 TAC---CATGCCCTTGCCCTCGCCCTAGCTATGAGTTGGCGCTGTG----- 1246
QY 168 ProLeuThrAspAlaLysThrArgSerHisAlaAspAspValGlyLeuValLeuSerSer 187
Db 1246 ----- 1246
QY 188 LeuGlnAlaSerIleTyrGlnGlyGlyTyrAsnArgPheAlaProTyr 203
Db 1247 -----GCGAGTTTATCCAGAGTGTGCTACAGCCGATTTGCCCTTAC 1288
RESULT 15
AAS93634
ID AAS93634 standard; cDNA; 2118 BP.
AC AAS93634;
XX
XX 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #29438.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
PI Dirmnac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR P-PSDB; ABG29447.
XX
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 29438; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2118 BP; 557 A; 516 C; 542 G; 503 T; 0 other;
Alignment Scores:
Pred. No.: 3,666-27 Length: 2118
Score: 405.50 Matches: 105
Percent Similarity: 37.38% Conservative: 9
Best Local Similarity: 34.43% Mismatches: 10
Query Match: 37.34% Indels: 181
DB: 23 Gaps: 7
US-09-809-545A-2 (1-203) x AAS93634 (1-2118)
QY 53 ValTyrThrSerAlaMetProGlyPheProTyrProAlaAlaThrAlaAlaTyr 72
Db 568 GTTGAATCAACTCAGTGCAGCGCTTCCGATCCAGAGCCACCGCGCGCTTAC 627
QY 73 ArgGlyAlaHisLeuAlaGlyArgGlyArgThrValTyrAsnThrPheArgAlaAla 92
Db 628 CGAGGGGGCACCCTGCAGAGCGCGGCTGCGACCGCTGTACACACCTTCAGGGCGCGG 687
QY 93 ProProProProIleProAlaTyrGlyVal----- 103
Db 688 CCCCCGCCCGATCCGAGCTCAGCGGCTGTGAAATCTGTAGACTTCCGAAATAT 747
QY 104 -----ValTyr--- 105
Db 748 TTGCTTTGTCTGATGATGACTTTAACAATATGCTGCTGCTGATCCGTGTATACT 807
QY 106 -----GlnGlu----- 107
Db 808 GAGACGACGAAACTAGCCTGTGCCACTGCCAGCTCAGCAGAACAAAGAGATTTC 867
QY 108 -----ProValTyrGly 111
Db 868 CACTAAAATGTTTGTGCATGCCAGTCCCTTTCAGCTTGTGGCGAGTC----- 921
QY 112 AsnTyrLeuLeuGlnGlyTyrAlaAlaTyrArgTyrAlaGlnProThrProAlaThr 131
Db 922 CCAAGGCTTGTGTAAGGTGTTATGCTGCATCCGCTACGCCAGCTACCCCTGCGACT 981
QY 132 AlaAlaAlaTyrSerAsp----- 137
Db 982 GCCGCTGCTACAGTACAGACAGACTGTCACTGATGTATGAGATGCGCTGAATATGAGG 1041
QY 137 ----- 137
Db 1042 AAGGAAAGAAATCATGCTGCTTCTGTTGCAGCAGATGAATTTCTGTACACCTTGCA 1101
QY 137 ----- 137
Db 1102 GGTGCTACCAAGTTACCAATTAGAAAGCAAGTTCAAGGGGTACAAAGAGTGTCAACAGC 1161
QY 137 ----- 137
Db 1162 GCCTTTCATTTTCTATCCCACTTTCACCTTACACATGCTGCAGTACATTAAACT 1221
QY 137 ----- 137
Db 1222 TTAGACAGTGGGATGCAATGCGACAAATATGACACTGCCAGCTTCATGAGATTTC 1281
QY 138 -----SerTyrGlyArgValTyrAlaAlaAspProTyrHisThr 151
Db 1282 TCGTTCCAGAGGGGTGAAGTTTACGAGAGATTTATGCTGCCAGCCCTTACACACCGCA 1341
QY 152 LeuAlaProAlaProThrTyrGlyValGlyValaMetAsnAlaPheAlaProLeuThrAsp 171
|||||

```

Db 1342 CTGCTCCAGCCCCCACTACGGCGTTGGTGCATGAT----- 1380
QY 172 AAlaYsThrArgSerHisAlaAspAspValGlyLeuValLeuSerSerLeuGlnAlaSer 191
Db 1381 -----AGACAGTAC-----AGTTCTTGTGTGTATCC 1407
QY 192 IleTyGInGlyGly 196
Db 1408 GTTGCACACTGTGTGA 1422

```

Search completed: March 16, 2003, 02:38:55
 Job time : 232 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: March 16, 2003, 02:29:52 ; Search time 1461 Seconds

(without alignments)
2250.297 Million cell updates/sec

Title: US-09-809-545A-2

Perfect score: 1086

Sequence: 1 MINKKAVNPYNGMKLNPVV.....VLSSLQASIVQGGYNRPAPY 203

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame.p2n.model -DEV=xip
-Q=cgnt_1/USPTO.spool/US09809545/rnac 07032003 153853 11467/app.query.fasta_1.391
-DB=EST -QPM=fastcap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blousum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pic -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US09809545 @CGN 1 1 2874 @runat 07032003 153853 11467 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAB -LARGEOUTRY -NEG_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:	
1: em_estba:	
2: em_esthum:	
3: em_estlin:	
4: em_estmu:	
5: em_estov:	
6: em_estpl:	
7: em_estro:	
8: em_hlc:	
9: gb_est1:	
10: gb_est2:	
11: gb_hlc:	
12: gb_est3:	
13: gb_est4:	
14: gb_est5:	
15: em_estfun:	
16: em_estom:	
17: gb_gss:	
18: em_gss_hum:	
19: em_gss_inv:	
20: em_gss_pln:	
21: em_gss_vit:	
22: em_gss_fun:	
23: em_gss_mam:	
24: em_gss_mus:	
25: em_gss_other:	
26: em_gss_pro:	
27: em_gss_rod:	

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	982.5	90.5	4039	11 BC026312	BC026312 Homo sapi
2	804.5	74.1	533	12 BF549922	BF549922 UI-R-E0-B
3	763.5	70.3	723	12 BG306387	BG306387 fms8d11.x
4	712	65.6	495	9 A1656926	A1656926 tct4e08.x
5	698.5	64.3	773	10 AV729057	AV729057 AV729057
6	691.5	63.7	773	10 AV729198	AV729198 AV729198
7	683	62.9	391	9 A1186272	A1186272 qd20b08.x
8	677.5	62.4	521	13 BM183349	BM183349 fv10e03.x
9	658.5	60.6	477	10 AM197589	AM197589 xma4g02.x
10	658.5	60.6	480	9 A1401040	A1401040 ch27a12.x
11	628.5	57.9	449	9 A1374891	A1374891 ch27a12.x
12	628.5	57.9	456	9 A1917298	A1917298 cs96g09.x
13	622.5	57.3	383	9 A1189407	A1189407 qd01h05.x
14	617	56.8	607	9 AUI45000	AUI45000 AUI45000
15	603	55.5	439	9 A1968176	A1968176 wu14a06.x
16	603	55.5	439	10 AM589795	AM589795 hg22d04.x
17	603	55.5	439	10 BE501380	BE501380 7a41n09.x
18	576.5	53.1	459	9 A1244212	A1244212 q18b11.x
19	569	52.4	430	9 A1655094	A1655094 w67a10.x
20	554.5	51.1	427	12 BF223478	BF223478 7g33e01.x
21	534.5	49.2	367	9 A1095813	A1095813 qb20g11.x
22	534	49.2	458	9 A1811011	A1811011 cu10c11.x
23	534	49.2	494	9 A1669300	A1669300 wb5b02.x
24	534	49.2	547	14 BM717439	BM717439 UI-E-EJ0-
25	534	49.2	555	14 BM714144	BM714144 UI-E-EJ0-
26	532.5	49.0	721	14 BM716595	BM716595 UI-E-DX1-
27	530	48.8	576	14 BM674802	BM674802 UI-E-EJ0-
28	530	48.8	610	14 BM727789	BM727789 UI-E-EJ0-
29	528.5	48.7	664	10 BB427515	BB427515 w85b02.x
30	525	48.3	1589	11 BC002124	BC002124 Mus muscu
31	523.5	48.2	556	10 AM015681	AM015681 UI-H-B10P
32	521	48.0	392	9 AA975235	AA975235 c936c08.b
33	521	48.0	346	9 A1674243	A1674243 wc45a12.x
34	517	47.6	594	12 BG380409	BG380409 UI-R-CT0-
35	516	47.5	679	14 BM727918	BM727918 UI-E-EJ0-
36	514	47.3	468	9 A1421028	A1421028 cf04d06.x
37	513	47.2	459	9 A1668768	A1668768 wc14a04.x
38	508.5	46.8	680	11 AK005186	AK005186 Mus muscu
39	506.5	46.6	372	9 A1799929	A1799929 wc41a05.x
40	500	46.0	773	13 B1856648	B1856648 603385804
41	495	45.6	690	14 BQ448092	BQ448092 UI-H-EU1-
42	481	44.3	495	9 A1682528	A1682528 wc54e02.x
43	478.5	44.1	443	9 AA773715	AA773715 af81c05.x
44	465	42.8	660	10 AM157266	AM157266 au93g01.x
45	462	42.5	525	9 A1390973	A1390973 mb94d11.y

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL
1	BC026312	BC026312	BC026312	BC026312.1	GI:20070932	Homo sapiens, clone IMAGE:4815500, mRNA.	4039 bp	HTC	linear	HTC 08-APR-2002
						Homo sapiens				
						Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;				
						Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
						1 (bases 1 to 4039)				
						Strausberg, R.				
						Direct Submission				
						Submitted (02-APR-2002) National Institutes of Health, Mammalian				

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk

Email: cgsabbs-romail.nih.gov
Tissue procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) and Shitaki
Toshiyuki and Piero Carninci (RIKEN)
cDNA library arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-ehgc.stanford.edu>
Contact: (Dickson, Mark) mcc@pxill.stanford.edu
Dickson, M., Schultz, J., Grimwood, J., Rodriguez, A., and Myers, J. M.

clone distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRaK Plate: 32 Row: 1 Column: 9
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA y1: 8922072
This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"  
/clone="IMAGE:4815500"  
/clone_type="Brain, hippocampus"  
/clone_tib="NIH MGC 95"
```

```

/Note="Vector: pBluescript"
BASE COUNT      1201 a      858 c      876 g     1104 t

```

Alignment Scores:

Pred. No.:	3,09e-92	403
Score:	982.50	189
Percent Similarity:	95.07%	
Best Local Similarity:	93.10%	
Query Match:	90.47%	
DB:	11	
	Gaps:	1

US-09-809-545A-2 (1-203) X BC026312 (1-4039)

Oy 1 MetThrAsnIysLysAlaValAsnProTyrThrAsnGlyTTPuLysLeuAsnProValVal
 Db 823 ATGACCAATAAAAGACCGCTCAACCTTATACAAATGGCTGGAATTGAAATCCAGTGTG

```
Oy      21 GYIAVAATyrSerProAspPheTyrAlaGlyThrValLeuLeuCysGlnIaaSerGln
          |||||::|||
Db      883 GGTCAGACTCAACAGTCCCGAATTCTATGCAGGCACGGTCCCTGTGTGCCAGCCCAACAG
```

Oy 41 GUGlySerSerMetTySerGlyProSerSerIeuValTyThrSeraIaMetProGly
 Db 943 GAGGATCTTCATGTACAGTGCCCCAGTTCACCTGTATATACTTCTGCATGCCAGGC

Qy 61 PheProTyrProAlaAlaThrAlaAlaAlaAlaTyrArgGlyAlaHisbLeuArgGlyArg
Db 1003 TTCCCGATCCAGCAGCACCGCGCGCGCGCTACCGAGGGCGCACCTGGAGGCGCGC

Qy 81 GlyArgThrValTyrAsnThrPheArgAlaAlaAlaProProProProIleProAlaTyr
Db 1063 GGTGGACCGTGACAAACCTTCAGGGCGGGGGCCCCCGATCCGGGCTAC

```

Oy 101 G1G1yValValTyrgIngluProValTyrg1yAsnLysLeuLeuGlnG1yG1yTyraLa
      |||||
Db 1123 GGGCGTGTtTTTACCAGATGcATTTATGcTcAGAc---ATTATGcGTtATGcCT
      |||||

```

Qy 121 AATyTArGyTyrAlaGlnProThrProAlaThrAlaAlaATySeraspSerTyGly
Db 1180 GCATACCGCTACGCCACGCTACCCCTGCACATGCGCTGCTACAGTGAAGTTACGGA

Qy 141 ArgValIYRAIaAaSPrOTyrHSHIStrLEuAlaPrOAlaPrOfHrtyrGlyVal 160
Dp 1240 CGAGTTTATGCTGCGACCCCTTACCAACAGCAGCTGTCCAGC-CGCACTACGGGTT 129

Qy 161 GlyAlaMetAsnAlaPheAlaProLeuThraspAlaIylsThraSerHisAlaIaasp 180
Dp 1299 GGCGCATGATGCTTTGCACTTTGACTGATGCCAAGACTAGAGCCATGCTGATGAT 1355

Dy 181 ValGlyLeuValIleuSerSerLeuGlnAlaSerIleTyrglngngylYtyrAsnArgPhe 200
Db 1359 GTGGGTCTCGTCTTTCTTCATTGCAGGCTACATATAACGAGGGGATCAACCGTTTT 1411

Qy	201	AlaProTyr	203
Db	1419	GCTCCATAC	1427

RESULT 2	533 bp	mRNA	linear	EST 12-DEC-2000
BF549922/c	BF549922			
LOCUS	UI-R-E0-bv-c-04-0-UI.r1	UI-R-E0	Rattus norvegicus	cDNA clone
DEFINITION	UI-R-E0-bv-c-04-0-UI 5'			sequence.

```

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db xref="taxon:10116"
/clone="UI-R-E0-bv-c-04-0-0-UI"
/clone_lib="UI-R-E0"
/dev_stage="embryonic"

```

BASE COUNT	134 a	125 c	164 g	109 t	1 others
------------	-------	-------	-------	-------	----------

Alignment Scores:	
Pred. No.:	9,72e-75
Score:	804.50
Percent Similarity:	94.08%
Best Local Similarity:	91.72%
Query Match:	74.08%
	Indels: 5
	Mismatches: 5
	Conservative: 4
	Matches: 15
	Length: 53

DB:	12	Gaps:	2
US-09-809-545A-2 (1-203) x BF549922 (1-533)			
Qy	39	AaenglnguglySerSerMetTyrserglProSerSerLeuValTyrThrSerAla---	57
Dd	532	AACCAAGAGGAGTATTCATGATGACAGTGGCCCAATTCACCTGTATATACCTTCGCAAT	473
Qy	58	-----MetProglYpHeProTyrProAlaAlaThrAlaAlaAlaAlaTyrArgly	74
Dd	472	TCTGTTCAGAGCCGCTTTCATATCCAGCCGCGCAGCTGCGACGTCGATACCGAGGG	413
Qy	75	AlaHisLeuAargGlyArgGlyArgThrValTyrAsnThrPheArgAlaAlaAlaAlaPro	94
Dd	412	GCTACCTTCAGAGCGCGGTGCGACCCGTGACCAACCTTAGCGTGGCGGCCCA	353
Qy	95	ProProlleProAlaTyrGlyGlyValValTyrGlnGluProValTyrGlyAsnYsleu	114
Dd	352	CCCCCAATCCCGGCTTAAGCGGCTTTTACCAAGATGATTTTATGTCGACAC---	296
Qy	115	LeuGlnGlyGlyTyrAlaAlaTyrArgTyrAlaGlnProThrProAlaThrAlaAla	134
Dd	295	ATTATGCTGTGCTTACGCTGATACCGATACGCCAGCCGACCCCTGCACGCTCTGCTCC	236
Qy	135	TyrTerAaPserTyrGlyArgValTyrAlaAlaAaPProTyrHisHisThrLeuAlaPro	154
Dd	235	TACAGTGCACGTTCGACGAGAGTTATATCTGCGCACCCCTTACACACACACTTCTCCA	176
Qy	155	AlaProTyrTyrGlyValGlyValaMetAaAlaPheAlaProLeuThrAspAlaYsThr	174
Dd	175	GGCCCAATACGGCGCTTGTGTCATGATGCTTTTGGCCCTTACCGATGCCAAGACT	116
Qy	175	ArgSerHisAlaAaPAspValGlyLeuValLeuSerSerLeuGlnAlaSerIlaTyrGln	194
Dd	115	AGGAGCCATGCTGATGATGCTGCTGCTCTTCTTCTTATTCAGCTAGTATATACCGA	56
Qy	195	GlyGlyTyrAsnArgPheAlaProTyr	203
Dd	55	GGGGGATACAAACCGTTTGTCTCCATAT	29
RESULT 3			
LOCUS 3	BG306387/c		
DEFINITION	BG306387	723 bp	linear
LOCATION	fm58d11.x1	zebrafish adult retina cDNA	Danio rerio cDNA clone
ACCESSION	U999493.3	similar to	TR:Q9UGW3 Q9UGW3 D41P2.2 ; mRNA sequence.
VERSION	BG306387		
KEYWORDS	EST.	GI:13103914	
SOURCE	zebrafish.		
ORGANISM	Danio rerio		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes		
AUTHORS	1 (bases 1 to 723)		
	Cyprinidae; Danio		
	Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy		
	S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood		
	,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,		
	Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,		
	Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.		
	and Wilson,R.		
	Washn Zebrafish EST Project 1998		
	Unpublished (1998)		
	Contact: Stephen L. Johnson		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: zbrafish@wustl.edu		
	Library constructed by: Chandra Tucker and Gregory Niemi DNA		
	Sequencing by: Washington University Genome Sequencing Center Clone		
	distribution: RessourcenzentrumPrimateBank, Berlin, Germany		
	(web address: www.tzpd.de)		
	Seq primer: 17 from Glibco		

FEATURES		High quality sequence stop: 472.	
source	Location/Qualifiers		
	1..723		
	/organism="Danio rerio"		
	/strain="wild-type"		
	/db_xref="taxon:7955"		
	/clone="4199493"		
	/clone_lib="zebrafish adult retina cDNA"		
	/sex="mixed"		
	/dev_stage="1-2 years"		
	/lab_host="E.Coli XL1-Blue MRF' (XL1-Blue MRF')"		
	/note=Vector: Lambda ZAP II (pBluescript SK-); Site-ECORI: Site 2: SalI: This zebrafish library was constructed by Dr. Susan E. Broekhoff (email: sbroekhoff@u.washington.edu) RZPD library number: 760"		
BASE COUNT	153 a 179 c 190 g 200 t		1 others
ORIGIN			
Alignment Scores:			
Pred. No.:	3,11e-70	Length:	723
Score:	763.50	Matches:	149
Percent Similarity:	77.94%	Conservative:	10
Best Local Similarity:	73.04%	Mismatches:	16
Query Match:	70.30%	Indels:	29
DB:	12	Gaps:	3
US-09-809-545A-2 (1-203) x BG306387 (1-723)			
Qy	1 MetThrAsnLysLeuAlaValaAsnProTyrThrAsnGlyTPrlySLeuAnPProValaVal	20	
Db	581 TTTCGAATTAATAACACGTCAACCCATATGACATGCGTGGAAAGTTGAATCAGTCCTG	522	
Qy	21 GlyAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuCysGlnAlaAsnGln	40	
Db	521 GGTGCAGTCTACAGCCAGAAATCTAT-----	493	
Qy	41 GluGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerLamEProGly	60	
Db	494 -----GCAATGCCAGGC	48	
Qy	61 PheProTyrProAlaAlaThrAlaAlaAla--AlaTyrArgGlyAlaHisLeuArgGly	79	
Db	482 TTCCCATACCCAGCAGCCACGGCGGACGTCGGCGTACAGAGGGGCACTTAAGAGA	422	
Qy	80 ArgGlyArgThrValTyrAsnThrPheArgAlaAlaAlaProProProProIleProAla	99	
Db	422 AGAGGCCGACCGTTTACACACGCTTTCGGGACAGCGCGCTCCCAACATCCAGCC	36	
Qy	100 TyrGlyGlyValValTyrGlnGluProValTyrGlyAsnLysLeuLeuGlnIlyGlyTyr	111	
Db	362 TATGGAGGTGGTTGTTTACACAGAGGGGTTTTACGGTGCAGAT---ATTATGTGGTTAC	300	
Qy	120 AlaAlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaTyrSerAspSerTyr	131	
Db	305 ACTGCCTACCGATACTACCTGCTGTAACGTACCGCCGCTGCTACAGTACAGTTAC	24	
Qy	140 GlyArgValTyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGly	15	
Db	245 GAGCGAGTTTAAAGTCGCGCAGCCCTTACACACAGCACTTCCTCCAGACGACATACAGC	18	
Qy	160 ValGlyAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAsp	17	
Db	185 GTTGGTGCATTAAGAGCTTTGGCAACCTTACATGATGCCAAGACCGAAGCCAGCCGAT	12	
Qy	180 AsnValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrGlnIlyGlyTyrAsnArg	19	
Db	125 GAGTGAGGTCGTTCTTTCTTTTGGAGGCTAAGTATATACGAGGTTGATACAGTCT	66	
Qy	200 PheAlaProTyr 203		
Db	65 TTGGCCCATAT 54		
RESULT 4			


```

Qy 101 GlyGlyValValTyrGlnGlnProValTyrGlyAsnLysLeuLeuGlnGlyGlyTyrAla 120
Db 123 GGGCGGTGTTGTTTACAGAGATGATTTATGTCAGAC---ATTATGTCGTTATGCT 179
Qy 121 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaTyrSerAspSerTyrGly 140
Db 180 GCATACCGCTACGCCACCTTACCCCTGCGCTGCTGCTACAGTACAGTACGGA 239
Qy 141 ArgValTyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGlyVal 160
Db 240 CGAGTTTATGTCGCGACCCCTACACACGACGACTTGTCCAGCCCACTACGCGCTT 299
Qy 161 GlyAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAspAsp 180
Db 300 GGTCCATGATGATGCTTTTGACACTTGATGCTGCAAGACTAGAGCCATGCTGATGAT 359
Qy 181 ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrGlnGlnGlyTyrAsnArgPhe 200
Db 360 GTGGGTCTCGTTCTTTCTTCAATGAGGCTATATATACGAGGGGGATACACCGTTT 419
Qy 201 AlaProTyr 203
Db 420 GCTCCATAC 428

RESULT 6
AV729198 773 bp mRNA linear EST 17-OCT-2000
LOCUS AV729198 HTC Homo sapiens cDNA clone HTCBCD09 5', mRNA sequence.
ACCESSION AV729198
VERSION AV729198.1 GI:10838619
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 773)
AUTHORS Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu
S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,W., Lu,G., Hu,R.,
Chen,J., Chen,Z. and Han,Z.
Homo sapiens cDNA HTC clones
JOURNAL Unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex. 45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
Source
1..773
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTCBCD09"
/clone_lib="HTC"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 184 a 232 c 185 g 170 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 1..3e-62 Length: 773
Score: 691.50 Matches: 131
Percent Similarity: 93.71% Conservative: 3
Best Local Similarity: 91.61% Mismatches: 8
Query Match: 63.67% Indels: 1
DB: 10 Gaps: 1

US-09-809-545A-2 (1-203) x AV729198 (1-773)

```

```

Qy 61 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 80
Db 3 TTCCCGTATCCAGAGCCAGCCGTCGCGCGCCCTACCGAGGGGGCGCACCTCGAGCCGC 62
Qy 81 GlyArgThrValTyrAsnThrPheArgAlaAlaAlaAlaProProProProlleProAlaTyr 100
Db 63 GTGCGACCGGTACCAACCTTCAGGGCGCGCGCGCCCGCCCGCCGATCCGCGCTAC 122
Qy 101 GlyGlyValValTyrGlnGlnProValTyrGlyAsnLysLeuLeuGlnGlnGlyTyrAla 120
Db 123 GGGCGGTGTTGTTTACAGAGATGATTTATGTCAGAC---ATTATGTCGTTATGCT 179
Qy 121 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaTyrSerAspSerTyrGly 140
Db 180 GCATACCGCTACGCCACCTTACCCCTGCGCTGCTGCTACAGTACAGTACGGA 239
Qy 141 ArgValTyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGlyVal 160
Db 240 CGAGTTTATGTCGCGACCCCTACACACGACCTTGTCTCCAGCCCACTACGCGCTT 299
Qy 161 GlyAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAspAsp 180
Db 300 GGTCCATGATGATGCTTTTGACACTTGATGCTGCAAGACTAGAGCCATGCTGATGAT 359
Qy 181 ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrGlnGlnGlyTyrAsnArgPhe 200
Db 360 GTGGGTCTCGTTCTTTCTTCAATGAGGCTATATATACGAGGGGGATACACCGTTT 419
Qy 201 AlaProTyr 203
Db 420 GCTCCATAC 428

RESULT 7
A1186273 391 bp mRNA linear EST 28-OCT-1998
LOCUS A1186273
DEFINITION q420b08.x1 Soares, placenta 8c09weeks 2bHPBct09W Homo sapiens cDNA
ACCESSION A1186273
VERSION A1186273.1 GI:3736911
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 391)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1800 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 369.

FEATURES
Source
1..391
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1724247"
/clone_lib="Soares placenta 8c09weeks 2bHPBct09W"
/dev_stage="Two placentae: One from 8 weeks and another
from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pUT73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGATGAGTGGAGCGCGCGCATTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pUT73 vector

```

(Pharmacia). Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 73 a 129 c 107 g 82 t

ORIGIN

Alignment Scores:

Pred. No.: 3.9e-62 Length: 391
Score: 683.00 Matches: 128
Percent Similarity: 99.23% Conservative: 1
Best Local Similarity: 98.46% Mismatches: 1
Query Match: 62.89% Indels: 0
DB: 9 Gaps: 0

US-09-809-545a-2 (1-203) x A1186273 (1-391)

QY 69 AAlaAlaATyArgTylAlaHisLeuArgGlyArgGlyArgThrValTyrAsnThrPhe 88
DB 1 GCGGCGGCTTACCGAGGGGCGACCTGCGAGGCGGCGGCTGCGACCTGTACACACCTTC 60
QY 89 ArgAlaAlaAlaProProProProProAlaTyrGlyValValTyrGlnGlnPro 108
DB 61 AGGCGCGGCGGCGGCGGCGGCGGCTTACCGAGGAGTGTGTATGATCAAGAGCTT 120
QY 109 ValTyrGlyAsnValLeuLeuGlnGlyTyrTyrAlaAlaTyrArgTyrAlaGlnProThr 128
DB 121 GTGTATGGCAATTAATGCTGTGAGGCTGTATGCTGATACCGCTACCGCCAGCCTTACC 180
QY 129 ProAlaThrAlaAlaAlaTyrSerAspSerTyrGlyArgValTyrAlaAlaAspProTyr 148
DB 181 CTGGCAGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
QY 149 HisHisThrLeuAlaProAlaProThrTyrGlyValGlyAlaMetAsnAlaPheAlaPro 168
DB 241 CACCAAGCACTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
QY 169 LeuThrAspAlaAlaAlaThrArgSerHisAlaAspAspValGlyLeuValLeuSerSerLeu 188
DB 301 TTGACTGAGTCCAGACACTGAGGCTGATGATGATGATGATGATGATGATGATGATGATG 360
QY 189 GlnAlaSerIleTyrGlnGlnGlyTyrAsn 198
DB 361 CAGGCTAGTATATACCGAGGGGATACAC 390

RESULT 8

BM183349/c 521 bp mRNA linear EST 26-JUL-2002
LOCUS fvl0e03.x2 zebrafish adult brain Dantio reio cDNA clone 5386085 3'
DEFINITION similar to TR:09UGW3 09UGW3 DV41P2.2 ; mRNA sequence.

ACCESSION BM183349
VERSION GI:17514307
KEYWORDS EST.
SOURCE zebrafish.

ORGANISM

Dantio reio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.

REFERENCE

AUTHORS 1 (bases 1 to 521)
Clark, M., Johnson, S. L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy
, S., Hillier, L., Kuebe, T., Martin, J., Beck, C., Wylie, T., Underwood
, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Peterson, B.,
Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Rittner, E.,
Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.
and Wilson, R.
TITLE Wabhu zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT Other ESTs: fvl0e03.y1
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@wustl.edu
CDNA Library Preparation: John Ngai. CDNA Library Arrayed by:

Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
ResourcenzenZentrumPrimateDatenbank, Berlin, Germany (web address:
www.rzpd.de)
Seq primer: -40UP
High quality sequence stop: 395.
Location/Qualifiers
1..521
/organism="Danio reio"
/db_xref="taxon:7955"
/clone="5386085"
/clone_id="zebrafish adult brain"
/sex="mixed male and female"
/tissue_type="brain"
/dev_stage="adult"
/lab_host="E. coli DH10B"
/note="Vector: pZIRPlox; Site_1: NotI; Site_2: SalI.
Original library was constructed in lambdaZIRPlox. Maas
excision of the cDNA library was performed to yield
pZIRPlox plasmids. Insert check was done in original
library."

FEATURES

source

BASE COUNT 111 a 124 c 149 g 137 t
ORIGIN
Alignment Scores:
Pred. No.: 2.22e-61 Length: 521
Score: 677.50 Matches: 135
Percent Similarity: 77.17% Conservative: 7
Best Local Similarity: 73.37% Mismatches: 13
Query Match: 62.38% Indels: 29
DB: 13 Gaps: 3

US-09-809-545a-2 (1-203) x BM183349 (1-521)

QY 21 GAlaAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuGlnAlaAsnGln 40
DB 520 GGTGAGTCTACAGCCCGAATTCAT----- 494
QY 41 GlnGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
DB 493 -----GCAGTGCAGGCG 482
QY 61 PheProTyrProAlaAlaThr--AlaAlaAlaAlaTyrArgGlyAlaHisLeuArgGly 79
DB 481 TTCCCATACCGACGACGACGCGGCGGCGGCTGACAGGCGGACACTTAAGAGA 422
QY 80 ArgGlyArgThrValTyrAsnThrPheArgAlaAlaAlaProProProProAla 99
DB 421 AGAGCGCGGACCGCTACACACGTTGCGGCGGCGGCGGCTGCGGCGGCGGCGGCGGCGG 362
QY 100 TrrGlyGlyValValTyrGlnGlnProValTyrGlyAsnValLeuGlnGlnGlyTyr 119
DB 361 TTTGAGGCTGTTTACCGAGCGGCTTTTACGCTGACAGT--ATTATGCTGTTAC 305
QY 120 AAlaAlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSerTyr 139
DB 304 ACTGCTACCGATACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 245
QY 140 GAlaArgValTyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGly 159
DB 244 GGACGAGTTTATGCTGCGGACCCCTACACGACGACCTTGTCCGACGACGACGACGACGAC 185
QY 160 ValGlyAlaMetAsnAlaAlaPheLeuThrAspAlaAlaValThrArgSerHisAlaAsp 179
DB 184 GTTGTGCTCATTAAGCTTTGACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 125
QY 180 AspValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrGlnGlyTyrAsnArg 199
DB 124 GATGTGGGTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 65

OY	200	Phealaprotyr 203 Db
	64	TTCGGCCCAIAT 53
RESULT 9		
LOCUS	AM197589	
DEFINITION	AM197589	477 bp mRNA linear EST 29-NOV-1999
ACCESSION	AM197589	xm14qg2.x1 NCI CGAP GC6 Homo sapiens cDNA clone IMAGE:2687050 3'
VERSION	AM197589	similar to TR:O43251 O43251 HYPOTHETICAL 39.5 KD PROTEIN.; mRNA sequence.
KEYWORDS	AM197589.1 GI:6476819	EST.
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgp. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	
JOURNAL	Unpublished (1997)	
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. cDNA library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: www.bio.linnl.gov/bbrp/image/image.html	
FEATURES		
source	Possible reversed clone: similarity on wrong strand Possible reversed clone: polyt not found Seq primer: -40UP from Gibco. Location/Qualifiers	
	1..477	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:2687090"	
	/clone_lib="NCI_CGAP_GC6"	
	/tissue_type="pooled germ cell tumors"	
	/lab_host="DH10B"	
	/note="Vector: pRTT3-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following Hsp purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonesids 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."	
BASE COUNT	100 a 145 c 123 g 109 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	1,96e-59	Length: 477
Score:	658.50	Matches: 125
Percent Similarity:	94.81%	Conservative: 3
Best Local Similarity:	92.59%	Mismatches: 6
Query Match:	60.64%	Indels: 1
Gaps:	10	Gaps: 1
US-09-809-S45A-2 (1-203) x AM197589 (1-477)		
OY	69	AlaalalaatYAAGGLyALahisieuAAGLyARGLyARThrValTYRanThrpe 88
Dd	1	GGCGGCCCTTAAGAGGGGCACAAGCAGAGCGCGGTGCGCACCGTAGAACACTTT 60
OY	89	ArglaaaialaProPProProllePfoAlaLYTGlyGYvalValTyrgLgrUpRo 108
		::::

D6		61	AGGCGCCGGGCGCCGCCGCCGCATCCCGCCTACAGCGCGGTGTTTATTACAGAGATGA	120
OY		109	VAllyrGtlyAenlySLeuleuEngInglyGIyTYrAlaAlaTyraGtyrAlaGlnProThr	128
D6		121	TTTTATGGTGAGAC---ATTATAGGGGTATATCTGCATACCCTACGCCAGCTTACC	177
OY		129	ProlaAThrrAlaAlaTyraSerAspSerGtyrGIyArVAlTYrAlaAlaAspProtyr	148
D6		178	CCTGCACATGCGGTGCTTCACGTACGTACGTAAACGACAGTTTATGCTGCCGACCCCTAC	237
OY		149	HishThrlleuAlaProAlaProThrTyrGIyValGIyalamelAsnAlaPheAlaPro	168
D6		238	CACCAACGACTGCTGCACGCCCCACCTACGGCGTTGGTGCATGAAGAAGCTTTGGACCT	297
OY		169	LeuthrAspaAlalystrArgSerHisAlaAspaAlaGIyleuValleuSerSerleu	188
D6		298	TTCAGTGTAGCCAGACTAGAGCACATGCTATATGATGGGTCTCGTTCTTCTCATTTG	357
OY		189	GlnAlaSerIleTyrgInglyGIyTYrAsnArghPheAlaProtyr	203
D6		358	CAGCTAGTATATACCGAGGGGATACACAGCTTTTGTCTCATAC	402
RESULT 10				
LOCUS		A1401040	480 bp mRNA linear EST 30-MAR-1999	
DEFINITION		t127a12.x1 NCI_CGAP_P228 Homo sapiens CDNA clone IMAGE:2119486 3,		
ACCESSION		A1401040		
VERSION		A1401040.1 GI:4244127		
KEYWORDS		EST.		
SOURCE		human.		
ORGANISM		Homo sapiens		
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE		NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
JOURNAL		National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
COMMENT		Tumor Gene Index		
		Unpublished (1997)		
		Contact: Robert Strausberg, Ph.D.		
		Email: cgapbs-femail.nih.gov		
		Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.		
		Emmer-Buck, M.D., Ph.D.		
		CDNA Library Preparation: M. Bento Soares, Ph.D.		
		CDNA library Arrayed by: Greg Lennon, Ph.D.		
		DNA Sequencing by: Washington University Genome Sequencing Center		
		clone distribution: NCI-CGAP clone distribution information can be		
		found through the I.M.A.G.E. Consortium/LNL at:		
		www.bio.linn.gov/bdrp/Image/Image.html		
		Insert Length: 460 Std Error: 0.00		
		Seq primer: -40Up from Gibco.		
FEATURES		Location/Qualifiers		
SOURCE		1..480		
		/organism="Homo sapiens"		
		/db_xref="taxon:9606"		
		/clone="IMAGE:2119486"		
		/clone_idb="NCI_CGAP_P228"		
		/sex="male"		
		/dev_stage="adult"		
		/lab_host="DH10B"		
		/note="Organ: prostate; Vector: pTZ19D-Pac (Pharmacia)		
		with a modified polylinker; Plasmid DNA from the		
		normalized library NCI CGAP P22 was prepared, and ss		
		circles were made in vitro. Following HAP purification,		
		this DNA was used as tracer in a subtractive hybridization		
		reaction. The driver was PCR-amplified cDNAs from a pool		
		of 5,000 clones made from the same library (clones		
		985608-986759, 1101192-1101959, and 1217928-1220615).		
		Subtraction by Bento Soares and M. Patricia Bonaldo."		
BASE COUNT		101 a 146 c 124 g 108 t		
ORIGIN		"		
Alignment Scores:				

www.bio.1nl.gov/bdrr/image/image.html
 Insert Length: 478 Std Error: 0.00
 Seq primer: -40UP from Gibco.

FEATURES

Location/Qualifiers
 1..456

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:229168"
 /clone_lib="NCI CGAP GC6"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA from the normalized library NCI CGAP GC4 was prepared, and 86 circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 97 a 137 c 115 g 107 t
 ORIGIN

Alignment Scores:

Pred. No.: 2,65e-56 Length: 456
 Score: 628.50 Matches: 119
 Percent Similarity: 94.57% Conservative: 3
 Best Local Similarity: 92.25% Mismatches: 6
 Query Match: 57.87% Indels: 1
 Gaps: 1

US-09-809-545a-2 (1-203) x A1917298 (1-456)

Qy 75 AAlaHisleuAargGlyAargGlyAargThyValTyraenthThreAargAlaAlaProPro 94
 Db 1 GCGCACCCTGAGGCGCGGTCGACCGCTGACACACCTTCAGGCGCGCGCGCGCGCG 60
 Qy 95 ProProleProAlaTyGlyGlyValValTyrgInGInProValTyGlyAAsnlyLeu 114
 Db 61 CCCCCGATCCCGGCTACGCGGCGTGTTCACAGATGATTTATGTCGACAGC--- 117
 Qy 115 LeuGInGlyTyGlyAlaAlaTyraGlyTyraGInProthProAlaThraAlaAla 134
 Db 118 ATTATGATGTTATGTCATACGCGTACCGCCGACCTTCAGCGACGTCGCTCC 177
 Qy 135 TyreerAspSerTyGlyAargValTyraAlaAspProTyRHisThleuAlaPro 154
 Db 178 TACAGTACAGTACGAGCAGATTATGTCGCCGCCCTACACACGACCTGCTCCA 237
 Qy 155 AlaProThTyGlyValGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 174
 Db 238 GCGCCCGCTACGCGGCTGTCGCTGCTTTCACCTTCATGATGACCAAGCT 297
 Qy 175 ArgSerHisAlaAspAspValGlyLeuValleuSerSerleuGInAlaSerIleTyGIn 194
 Db 298 AGGAGCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 357
 Qy 195 GlyGlyTyraAlaArgPheAlaProTyR 203
 Db 358 GGGGATACCAACCGTTTTCCTCATAC 384

RESULT 13

LOCUS A1189407 383 bp mRNA linear EST 28-OCT-1998
 DEFINITION g001h05.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
 IMAGE:1722489 3', mRNA sequence.
 ACCESSION A1189407
 VERSION A1189407.1 GI:3740616
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 383)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-rc@mail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.1nl.gov) for further information.
 Insert Length: 486 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 381.

FEATURES

source

Location/Qualifiers
 1..383

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1722489"
 /clone_lib="Soares_pregnant_uterus_NbHPU"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: pT73-Pac; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-ACTGGAAGATTCGCGCGCCCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."

BASE COUNT 69 a 122 c 102 g 90 t
 ORIGIN

Alignment Scores:

Pred. No.: 8.88e-56 Length: 383
 Score: 622.50 Matches: 118
 Percent Similarity: 94.53% Conservative: 3
 Best Local Similarity: 92.19% Mismatches: 6
 Query Match: 57.32% Indels: 1
 Gaps: 1

US-09-809-545a-2 (1-203) x A1189407 (1-383)

Qy 72 TyraGlyAlaHisleuAargGlyAargGlyAargThyValTyraenthThreAargAlaAla 91
 Db 1 TACGAGGGGGGCGACCTGCGAGGCGCGGTCGACCGGTACACACCTTCAGGCGCGCG 60
 Qy 92 AlaProProProleProAlaTyGlyGlyValValTyrgInGInProValTyGly 111
 Db 61 GCGCCCGCGCGCGCGCTACGCGCGGCTGTTTACAGATGATTTATGCT 120
 Qy 112 AsnlyLeuLeuGInGlyTyGlyAlaAlaTyraGlyTyraGInProthProAlaThr 131
 Db 121 GCAGAC--ATTATGATGTTATGTCATACCGCTACGCGCGCCTTCGCACT 177
 Qy 132 AlaAlaAlaTySerAspSerTyGlyAargValTyraAlaAspProTyRHisThleu 151
 Db 178 GCGCGCTGCTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 237
 Qy 152 LeuAlaProAlaProThTyGlyValGlyAlaAlaAlaAlaAlaAlaAlaAlaAla 171
 Db 238 CTGCTCAGCCCGCCACCTACGCGCGGCTGTCATGATGCTTTGCACTTGATGAT 297
 Qy 172 AlaAlaAlaTySerHisAlaAspAspValGlyLeuValleuSerSerleuGInAlaSer 191
 Db 298 GCGAAGACTAGACCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 357
 Qy 192 IleTyrgInGlyTyGlyAAsnArg 199
 Db 358 ATATACGAGGGGATACACCGT 381

RESULT 14

AU145000/c

```

LOCUS       AUI45000                      607 bp    mRNA    linear    EST 05-AUG-2002
DEFINITION  AUI45000 HEMBA1 Homo sapiens cDNA clone HEMBA1003591 3', mRNA
ACCESSION   AUI45000
VERSION     AUI45000.1  GI:11006521
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE   1 (bases 1 to 607)
AUTHORS     Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Salto,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura,Y., Nagai,T., Sugano
            S. and Isegai,T.).
            HRI human cDNA project (Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S.,
            Salto,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura
            Y., Nagai,T., Sugano,S., Isegai,T.)
            Unpublished (2000)
JOURNAL     Genomics Laboratory
COMMENT     Contact: Takao Isegai
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
            Research Institute; cDNA library construction; Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute
FEATURES             Location/Qualifiers
     source          1..607
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HEMBA1003591"
                     /clone_lib="HEMBA1"
                     /tissue_type="whole embryo, mainly head"
                     /dev_stage="embryo, 10 weeks"
                     /note="Vector: pME18SFL3"
BASE COUNT        140 a 152 c 179 g 129 t 7 others
ORIGIN
1
Alignment Scores:
Pred. No.:        6,47e-55      Length:      607
Score:            617.00        Matches:    136
Percent Similarity: 83.04%      Conservative: 6
Best Local Similarity: 79.53%   Mismatches: 21
Query Match:      56.81%       Indels:      8
DB:               9            Gaps:        2
US-09-809-545A-2 (1-203) x AUI45000 (1-607)
QY      40 GINGUGIYSerSerMetTySerGly---ProSerSerLeuValTyThrSerAla-Me 58
DB      600 CAGGAGGGATTTTTCATCGACAGNGCCCCCGACGTTCACTTGGAAANATTTTNGCAANT 541
QY      58 cProGlyPhePro-TyPProAlaAlaThraAlaAlaAlaAlaTyArgGlyAlaHisIleuA 78
DB      540 GCCAGGCTTCCGATTCACAGAGCCACCGCGCGCGCTTCCAGAGGGCCCACTTGC 481
QY      78 TGGIYArg-GIYArgThraValTyraenThrPheArgAlaAlaAla-ProProProProIi 97
DB      480 GAGNCGCGGGGTGCACCGGTGACCAACACTTCAGGGCGCGCGCGCGCGCGCGAT 421
QY      97 ePro-AlaTyArgIy-GlyValValTyGInGluProValTyArgIyAsnIySleuLeuGIn 116
DB      420 CCGCGGCGCTACCGCGCGGTGTGTTTACAGAGATGATTTTATGTGCAGAC--ATTAT 364
QY      117 GIGYIYTyAlaAlaAlaTyArgTyArgIAGInProThProAlaThraAlaAlaTySer 136
DB      363 GGGGGTTAGCTGCATACCGCTACGCCACCTTACCCCTGCCCTGCGCTACAGT 304
QY      137 AaSPSerTyGIYArgValTyArgAlaAlaAaSPProTyHisIsthraAlaProAlaPro 156

```

```

DB      303 GACAGTTACGACGAGCTTATGCTGCCGACCCCTTACACACGACGACTTGCCTCCAGCCCC 244
QY      157 ThTYrGIYValGIYAlaMeCasAlaAlaPheAlaProLeuThraSPAlaIySerThraYrSer 176
DB      243 ACCTACGGCGCTTGGTGGCCATGATGATCTTTTGACACTTTTGACATGATGCCAAGATGAGAC 184
QY      177 HisAlaAaSPaSPValGIYLeuValIeuSerSerLeuGInAlaSerIleTyGIInGlyIy 196
DB      183 CATGCTGATGATGTGGGTCTGCTTCTTTCATGTCAGGCGCTAGTATATACGAGGGGGA 124
QY      197 TYraenArGPheAlaProTy 203
DB      123 TACAACCGTTTGCTTCATAC 103
RESULT 15
LOCUS     A1968176                      439 bp    mRNA    linear    EST 25-AUG-1999
DEFINITION  wu14806.x1 NCI CGAP GC6 Homo sapiens cDNA clone IMAGE:2516914 3',
            similar to TR:O43251 O43251 HYPOTHETICAL 35.5 KD PROTEIN. ; mRNA
            sequence.
ACCESSION   A1968176
VERSION     A1968176.1  GI:5764994
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE   1 (bases 1 to 439)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cga@bcr-rt@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmerit-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/ILMIL at:
            www.bio.livl.gov/bdrip/image/image.html
            Seq primer: -40UP from Glibco.
FEATURES             Location/Qualifiers
     source          1..439
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2516914"
                     /clone_lib="NCI CGAP GC6"
                     /tissue_type="Pooled germ cell tumors"
                     /lab_host="DH10B"
                     /note="vector: pRTT3D-Pac (Pharmacia) with a modified
                     polylinker. Site_1: Not I; Site_2: Eco RI; Plasmid DNA
                     from the normalized library NCI-CGAP GC4 was prepared, and
                     ss circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (clonoids
                     1257096-1258631, 1469064-1470983, and 1475592-1476743).
                     Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT        91 a 142 c 114 g 92 t
ORIGIN
1
Alignment Scores:
Pred. No.:        1,22e-53      Length:      439
Score:            603.00        Matches:    118
Percent Similarity: 88.15%      Conservative: 1
Best Local Similarity: 87.41%   Mismatches: 2
Query Match:      55.52%       Indels:     15
DB:               9            Gaps:        1
US-09-809-545A-2 (1-203) x A1968176 (1-439)

```

```
Qy 69 AlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArgGlyArgThrValTyrAsnThrPhe 88
    |||
    |||
    |||
Db 1 GCGGCCGCTACCGAGGGCGCACCTCGAGGCGCGGTCCGACCGGTACACACCTTC 60
    |||
    |||
    |||
Qy 89 ArgAlaAlaAlaProProProProProIleProAlaTyrGlyGlyValValTyrGlnGluPro 108
    |||
    |||
    |||
Db 61 AGGCGCGCGCGCGCGCGCGCGCGATCCCGGCTACGGCGG----- 101
    |||
    |||
    |||
Qy 109 ValTyrGlyAsnLysLeuLeuGlnGlyGlyTyrAlaAlaTyrArgTyrAlaGlnProThr 128
    |||
    |||
    |||
Db 102 -----GGTGTATGCTGCATACCGCTACGCCAGCCTACC 137
    |||
    |||
    |||
Qy 129 ProAlaThrAlaAlaAlaTyrSerAspSerTyrGlyArgValTyrAlaAlaAspProTyr 148
    |||
    |||
    |||
Db 138 CTGCGCACTGCGCGTGCCTACAGTGACAGTTACGAGCGATTATGCTGCCGACCCCTAC 197
    |||
    |||
    |||
Qy 149 HisHisThrLeuAlaProAlaProThrTyrGlyValGlyAlaMetAsnAlaPheAlaPro 168
    |||
    |||
    |||
Db 198 CACCACGCACTTGCTCCAGCCCCCACCTACGCGCTTGCGCATGATGCTTTGCACCT 257
    |||
    |||
    |||
Qy 169 LeuThrAspAlaLysThrArgSerHisAlaAspAspValGlyLeuValLeuSerSerLeu 188
    |||
    |||
    |||
Db 258 TTGACTGATGCCAAGACTAGAGGCCATGCTGATGATGAGGCTCTGTTCTTTCATTG 317
    |||
    |||
    |||
Qy 189 GlnAlaSerIleTyrGlnGlyGlyTyrAsnArgPheAlaProTyr 203
    |||
    |||
    |||
Db 318 CAGGCTAGTATATACCGAGGGGATACACCGTTTGTCTCCATAC 362
    |||
    |||
    |||
```

Search completed: March 16, 2003, 04:00:24
Job time : 1466 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 16, 2003, 02:32:41 ; Search time 76 Seconds
(without alignments)
819.150 Million cell updates/sec

Title: US-09-809-545A-2

Perfect score: 1086

Sequence: 1 MTNKKAVNPTNMGKLNPPV.....VLSSLOASTYQGGYNRPAPY 203

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 segs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=x1p
-Q=/cgn2_1/USPTO.spool/US09809545/rnat.07032003.153853.11480/app_query.fasta_1.391
-DB=Issued_Patents_NA -QPM=fastacp -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsun62 -TRANS=human40.cdl
-LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USPR=US09809545 @CGN 1.1 61 @rnat.07032003.153853.11480 -NCPU=6 -ICPU=3
-NO_XLXRY -NO_MMAT -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -YGAPEXT=0.5 -Fgapop=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELEXT=7

Database :

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	955	87.9	2372	4	US-09-145-391-1
2	467.5	43.0	1506	3	US-09-176-657-5
3	101	9.3	914	2	US-08-935-450-10
4	98.5	9.1	821	4	US-08-990-823-62
5	96	8.8	3231	1	US-08-195-152-1
6	94	8.7	2368	4	US-09-343-011B-3
7	94	8.7	43280	2	US-08-804-227C-1
8	92.5	8.5	2371	2	US-08-343-443B-1
9	92.5	8.5	2412	1	US-08-437-027-18
10	92.5	8.5	28958	1	US-08-258-261B-6
11	92.5	8.5	28958	1	US-08-456-837-6
12	92.5	8.5	28958	1	US-08-457-342-6

13	92.5	8.5	28958	1	US-08-457-645A-6	Sequence 6, Appl1
14	92.5	8.5	28958	1	US-08-458-076A-6	Sequence 6, Appl1
15	92.5	8.5	28958	1	US-08-764-233A-4	Sequence 6, Appl1
16	92.5	8.5	28958	1	US-08-457-335A-6	Sequence 6, Appl1
17	92.5	8.5	28958	1	US-08-729-214-6	Sequence 6, Appl1
18	92.5	8.5	28958	3	US-09-028-934-6	Sequence 6, Appl1
19	92.5	8.5	49377	1	US-08-764-233A-1	Sequence 1, Appl1
20	92	8.5	53526	3	US-08-658-136-2	Sequence 2, Appl1
21	92	8.5	53577	3	US-08-658-136-1	Sequence 1, Appl1
22	91	8.4	1896	4	US-09-343-011B-4	Sequence 4, Appl1
23	90.5	8.3	10095	3	US-08-822-586-45	Sequence 45, Appl1
24	89	8.2	1140	3	US-09-023-173-4	Sequence 4, Appl1
25	88	8.1	2923	6	5187076-5	Patent No. 5187076
26	87.5	8.1	3155	4	US-09-442-100-7	Sequence 7, Appl1
27	87.5	8.1	44377	2	US-08-804-227C-7	Sequence 7, Appl1
28	87.5	8.1	44377	2	US-08-804-198-1	Sequence 1, Appl1
29	87	8.0	1352	1	US-08-552-142A-10	Sequence 10, Appl1
30	87	8.0	1452	1	US-08-552-142A-16	Sequence 16, Appl1
31	87	8.0	2418	1	US-08-462-184-1	Sequence 1, Appl1
32	87	8.0	2418	1	US-08-441-147-1	Sequence 1, Appl1
33	87	8.0	2418	5	PCT-US95-07536-1	Sequence 1, Appl1
34	86.5	8.0	2040	2	US-08-031-538-10	Sequence 10, Appl1
35	86.5	8.0	2888	4	US-08-765-907A-1	Sequence 1, Appl1
36	86.5	8.0	13807	4	US-09-052-469-5	Sequence 5, Appl1
37	86.5	8.0	14060	3	US-08-658-136-4	Sequence 4, Appl1
38	86.5	8.0	14148	4	US-09-052-469-7	Sequence 7, Appl1
39	86	7.9	735	4	US-09-336-536-47	Sequence 47, Appl1
40	86	7.9	1801	4	US-09-336-536-46	Sequence 46, Appl1
41	85	7.8	654	4	US-08-998-416-1144	Sequence 1144, Ap
42	85	7.8	1494	4	US-09-255-502-1	Sequence 1, Appl1
43	85	7.8	1497	1	US-08-322-677A-6	Sequence 6, Appl1
44	85	7.8	1497	1	US-08-322-676-6	Sequence 6, Appl1
45	85	7.8	1497	3	US-08-898-218-6	Sequence 6, Appl1

ALIGNMENTS

RESULT 1
US-09-145-391-1
Sequence 1, Application US/09145391
Patent No. 6194171
GENERAL INFORMATION:
APPLICANT: Pulst, Stefan M.
APPLICANT: Shibata, Hiroki
TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
TITLE OF INVENTION: Products Related Thereto and Methods of Using Same
FILE REFERENCE: CE 3093
CURRENT APPLICATION NUMBER: US/09/145,391
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 2372
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (987)..(1979)
US-09-145-391-1

Alignment Scores:

Pred. No.: 1.3e-85 Length: 2372
Score: 955.00 Matches: 189
Percent Similarity: 86.94% Conservative: 4
Best Local Similarity: 85.14% Mismatches: 9
Query Match: 87.94% Indels: 21
DB: 4 Gaps: 2

US-09-809-545A-2 (1-203) x US-09-145-391-1 (1-2372)

Qy 1 MetThrIasnLysLysGlnAlaValAsnProTyrTrhAsnGlyTTrpLysIleuAsnProValVal 20
Db 1572 ATGACAAATATAAAGACCGCTTATACCAATGCGTGAATGGAATCCAGTTGTG 1631

Score: 101.00 Matches: 48
 Percent Similarity: 37.06% Conservative: 15
 Best Local Similarity: 28.24% Mismatches: 59
 Query Match: 9.30% Indels: 48
 DB: 2 Gaps: 7

US-09-809-545A-2 (1-203) x US-08-935-450-10 (1-914)

```

Qy 24 TyserProAepheTyraGlyThrValLeuLeuCySGlnAlaAengIngluGlySer 43
    |||
Db 230 TACGACCTT-----GCTCGAACCCCCACAGGGGCC 259
Qy 44 SerMetTyserGlyProSerSerLeuValTyThrSerAla-MetProGlyPheProTy 63
    |||
Db 260 AGCACCTACATAGAACAGACGACATCCCTGCTACAGCCCAATACAGACACCCACC 319
Qy 63 rProAlaAlaThrAla-----AlaAlaAlaTyraGlyAlaHisLeuAlaGlyAlaGly 81
    |||
Db 320 GTCCAGCAGTACAGCCCTTCCACAGCCGAGTTACAGCCAGCCACCTTAC--AACACAGG 376
Qy 81 YargThrValTyraSerThrPheArgAlaAlaAla-ProProProProIleProAlaTyrg 101
    |||
Db 377 GAGGTACAGCCAGGAGTTACAGAGCCACCGCTCCCTCCACACACACCTGCTTACA 436
Qy 101 lYglValValTyrgIngluProValTyrgIYasnIYsLeuLeuInglYglTYraAla 121
    |||
Db 437 AC-----TATGGAGC-----TACGGCGTTACACAC 463
Qy 121 lATyraGTYraAlaGlnProThrProAlaThrAlaAlaAlaTyserAspSerTyrgIY 141
    |||
Db 464 CGGCCCTTATACCCACAGCCGCCACCCCGCCAGCAGACCTTAC----- 507
Qy 141 rGValTYraAlaAlaAspProTyRHisThrLeuAlaProAlaProThrTyrgIYValG 161
    |||
Db 508 -----CTCAGCCCAACTATATACCACT 529
Qy 161 lYAlaMetAsnAlaPheAlaProLeuThrAspAlaAlaYsThrArgSerHisAlaAsp 181
    |||
Db 530 ATCAGCACTATAGCCAGCGATGGAGACACTACTACTAGAACACAGCCGCGCT 579
Qy 181 aIGlyLeuValLeuSerSerLeuGln 189
    |||
Db 580 --GGCGCATCTACTACGGAGACTACGA 603

```

RESULT 4

US-08-990-823-62
 ; Sequence 62, Application US/08990823D
 ; Patent No. 6228371
 ; GENERAL INFORMATION:
 ; APPLICANT: Nano, Francis
 ; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
 ; TITLE OF INVENTION: Immunostimulatory Peptides
 ; FILE REFERENCE: 49086
 ; CURRENT APPLICATION NUMBER: US/08/990, 823D
 ; CURRENT FILING DATE: 1997-12-15
 ; EARLIER APPLICATION NUMBER: US 96/10375
 ; EARLIER FILING DATE: 1996-06-14
 ; EARLIER APPLICATION NUMBER: 60/000,254
 ; EARLIER FILING DATE: 1995-06-15
 ; NUMBER OF SEQ ID NOS: 113
 ; SOFTWARE: PatentIN Ver. 2.0
 ; SEQ ID NO 62
 ; LENGTH: 821
 ; TYPE: DNA
 ; ORGANISM: Mycobacterium tuberculosis
 ; FEATURE:
 ; NAME/KEY: Modified base
 ; OTHER INFORMATION: n represents a or g or c or t/u
 US-08-990-823-62

Alignment Scores: 0.348 Length: 821
 Pred. No.: 98.50 Matches: 59
 Score:

Percent Similarity: 28.46% Conservative: 15
 Best Local Similarity: 22.69% Mismatches: 95
 Query Match: 9.07% Indels: 91
 DB: 4 Gaps: 9

US-09-809-545A-2 (1-203) x US-08-990-823-62 (1-821)

```

Qy 1 MetThrAsnIYsYsAlaValaAlaAsnProTyThrAsnGlyTYrIYsLeuAsnProVala 20
    |||
Db 70 ATGGCCAGCAGCGTGTGCGCCAGCCGACCCCAATCCCGCATGTGCAACCGGTTCG 129
Qy 21 GlYAlaValTYrSerProAepheTyraGlyThrValLeuLeuCySGlnAlaAengIn 40
    |||
Db 130 GGCAGCGCGTTCGAGCCGAGCAGCAGCGTGGCGGTATCAGTCCGTGCGCTTCAACCCG 189
Qy 41 GluIYserSerMetTyserGlyProSer-----SerLeuValTy 54
    |||
Db 190 AGGCGCTGGGCGAGAGACTCAAGTCCAGCCCGGTGTCTCCAAACCCGCGTGTCAAC 249
Qy 55 ThrSerAlaMetProGlyPheProTyProAlaAlaThrAlaAlaAlaTYraGly 74
    |||
Db 250 TCGATGTGTGCGCCCAACCAACCGCCGCTACACGACTCCGCGGCGACCGCGGA 309
Qy 74 ----- 74
Db 310 GGAAGGCGCCGNTCGGATNCAACGAGGTTCNACCGCGCGCTGCTTCNGATTGA 369
Qy 75 -----AlaHisLeuArgIYArg-----GlyArgThrValTy 85
    |||
Db 370 CCGGCGACGTAACCCCGGTATGGCAGCTACGGGAGAACAACTGGCCGCGCACGGCCAC 429
Qy 86 AsnThrPheArgAlaAlaAlaProProProProIleProAlaTyrgIYValValTy 105
    |||
Db 430 CTCGGCGTGTACAGTTACCGCCCGAGCCGAGCCGCGCGCTGTGT----- 480
Qy 106 GluInlProValTYrgIYasnIYs-----LeuLeuInglYglY 118
    |||
Db 481 -----GGTTCGCGCGCGCGCATGTGTCTTCAAGAGAGACCGCGATTTCATCTA 534
Qy 119 -----TyraAlaAlaTYraTYraGTYraAlaGlnProThrProAlaThr 131
    |||
Db 535 CGGCCAGTCCCTGAAATCTGACGTGGGGCGTACCGGCCGCGGACCGCGCATTCACCACT 594
Qy 132 AlaAlaAlaTYrSerAspSerTy----- 139
    |||
Db 595 GGGCGAGTATTTCCGATCGACATCGGACCGCAACCGCGTGGCCCATTCGCGTTCC 654
Qy 140 -----GlyArgValTYraAlaAspProTyRHis----- 150
    |||
Db 655 GCTGGCTGGCGCGCGCGCGGAGCGCGCATGTGCGCG-----CATTTGCGC 699
Qy 151 -----ThrLeuAlaProAlaProThrTyrgIY 159
    |||
Db 700 CTATGACCCGCACTGAGCCCTGAGCAATGTTGCGCTTCAACCCCGCGGTTCGCGT 759
Qy 160 ValGlYAlaMetAsnAlaPheAlaProLeuThrAspAlaAlaYsThrArgSerHisAla 179
    |||
Db 760 GCTGGAATCTCTGACAGCGGTGTATCGGTACGACACCGGTGTGTGATGACATCGCAC 819

```

RESULT 5

US-08-195-152-1
 ; Sequence 1, Application US/08195152
 ; Patent No. 5679541
 ; GENERAL INFORMATION:
 ; APPLICANT: Bonini, Nancy M.
 ; APPLICANT: Leiserson, William M.
 ; APPLICANT: Benzer, Seymour
 ; TITLE OF INVENTION: PROGRAMMED CELL DEATH ANTAGONIST
 ; TITLE OF INVENTION: PROTEINS
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Flehr, Hohnbach, Teest, Albritton & Herbert
 ; STREET: 4 Embardadero Center, Suite 3400

CITY: San Francisco
 STATE: California
 COUNTRY: United States
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/195,152
 FILING DATE: 14-FEB-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Treccatin, Richard F.
 REGISTRATION NUMBER: 31,801
 REFERENCE/DOCKET NUMBER: A-59551/RFT/RMS
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELEX: 910 277299
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3231 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: unknown
 US-08-195-152-1

Alignment Scores:
 Pred. No.: 3.93 Length: 3231
 Score: 96.00 Matches: 45
 Percent Similarity: 37.57% Conservative: 23
 Best Local Similarity: 24.86% Mismatches: 63
 Query Match: 8.84% Indels: 51
 DB: Gaps: 6

US-09-809-545a-2 (1-203) x US-08-195-152-1 (1-3231)
 QY 42 GAGSerserMetTy-----SerGlyProserSerLeu 52
 DB 1062 GGATCCAAATTGTGACCGCTGCGACCTGCGCCAGCAATCCCGCTGAGCAGAGAGAGAGAGCGG 1121
 QY 53 ValTYrThSerAlaMetProGlyPheProTYrProAlaAlaThrAlaAlaAlaTYr 72
 DB 1122 GTCACACTCTTCGCGACATG-----GCAAGCGGACAGAGCGGCTCTAC 1163
 QY 73 ArgGlyAlaHisLeuArgGlyArgGlyArgThrValTYrAsnThrPheArgAlaAlaAla 92
 DB 1164 GACGGCAACAT-----GACTACTACTACTACACACGATGACGACGATACAG 1211
 QY 93 ProProProProIleProAlaTYrGlyGlyValValTYrGlnGluProValTYrGlyAsn 112
 DB 1212 CCGCGCCCTTCTACTCCGGATCGGAATCCTTATCGCGCGCAACGCGCGACGCGAG 1271
 QY 113 LysLeuLeuGlnGlyGlyTYrAlaAlaTYrArgTYrAlaGlnProThrProAlaThrAla 132
 DB 1272 GCCAAGATGGAACCCGAGGCGGACGCTGCGCGGCTGCTTACGCGCCAGCTATGCC 1331
 QY 133 AlaAla----- 134
 DB 1332 GCCAGCGGCAACAACAATCGCAGCTGTACAGAGTCCGTAACCGCGCTACACAACATTC 1391
 QY 135 -----TyrSerAspSerTYrGlyValTYrAlaAlaAsp 146
 DB 1392 GGGCAGCAGAGACTACGGCGGCTACTACACAGACAGTACGCACTATTACAGTCCGGCC 1451
 QY 147 ProTYrHisIleThrLeuAlaProAlaPro-----ThrTYrGlyValGly--- 161
 DB 1452 AACTACTCAACCGATGCGGCTACGCTGCGCCAGCTCGAGTGCAGTATGAGCACTGCTTC 1511
 QY 162 AlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAlaAspAspVal 181

DB 1512 CATGTGGCGGCTCTCGAATCTTCCAGAGTCCCGAGGACACCCACTC-GACGAGCC 1570
 QY 182 Gly 182
 DB 1571 GGT 1573

RESULT 6
 US-09-343-011B-3
 ; Sequence 3, Application US/09343011B
 ; Patent No. 6300473
 ; GENERAL INFORMATION:
 ; APPLICANT: Stephane Richard
 ; TITLE OF INVENTION: SLM-1 AND SLM-2, NOVEL
 ; FILE REFERENCE: A32561
 ; CURRENT APPLICATION NUMBER: US/09/343,011B
 ; PRIOR FILING DATE: 2001-05-11
 ; PRIOR APPLICATION NUMBER: CA 2265271
 ; NUMBER OF SEQ ID NOS: 9
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 3
 ; LENGTH: 2368
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 US-09-343-011B-3

Alignment Scores:
 Pred. No.: 4.08 Length: 2368
 Score: 94.00 Matches: 54
 Percent Similarity: 36.93% Conservative: 11
 Best Local Similarity: 30.68% Mismatches: 51
 Query Match: 8.66% Indels: 60
 DB: Gaps: 12

US-09-809-545a-2 (1-203) x US-09-343-011B-3 (1-2368)
 QY 73 ArgGlyAlaHisLeuArgGlyArgThrValTYrAsnThrPhe----- 88
 DB 1673 CGGGCCCGAGATATGAGGACGAGGATCAGAAATTAATCCCAAGCTTCATCAAGGGCC 1732
 QY 89 ArgAlaAlaAlaProProProProIleProAlaTYrGlyGlyValValTYrGlnGluPro 108
 DB 1733 CGTGGCGGTCTGTTCCACCACCAACCACTGACGAGGAGTGTCTTACCCCTCGG-- 1789
 QY 109 ValTYrGlyAsnLysLeuLeuGlnGly----- 117
 DB 1790 -----GGGACCACTGTACCCGCTGAGCTCTCCAGTCCCAATGACAGAGTTC 1843
 QY 118 -----GlyTYrAlaAla-----TyrArgTYrAlaGlnProThr 128
 DB 1844 CCAACACTCGAGCCCGGCGGACGAGATACCAAGATACAGA---GCACCCCACT 1900
 QY 129 ProAlaThrAlaAla-----AlaTYrSerAspSerTYrGlyArgValTYrAla 144
 DB 1901 CCAAGCTCATGATGTATGAGAAATATGGATATGATGCTATGGGGTGAATATGAT 1960
 QY 145 AlaAspProTYr-----HisIleThrLeuAlaProAlaProThr----- 157
 DB 1961 GACCAAGCTATGAGCTTATATATAGTACGACCCCAACAACAAGTGTGCTGAA 2020
 QY 158 -----TYrGlyValGly-----AlaMetAsnAlaPheAlaProLeuThr 170
 DB 2021 TACTATGACTACGATGAGTAAACGAGATGCTTACGACAGCTACCAACAGAGAA 2080
 QY 171 AspAlaLysThrArgSerHisAlaAspAspValGlyLeuValLeuSerSerLeuGlnAla 190
 DB 2081 TGGGCCAACAATCGC-----TTCAGCTCTGAAGGCA 2110
 QY 191 -----SerIleTYrGlnGlyGlyTYrAsnArgPheAlaProTYr 203
 DB 2111 CCACCAACCAAGGTCAAGCCAGAGGGGATTC---AGGAGACACCCCTAT 2155

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00494
FILING DATE: 19-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/06123
FILING DATE: 20-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 989.6121P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2371 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 25..1992
US-08-343-443B-1

Alignment Scores:
Pred. No.: 5.75 Length: 2371
Score: 92.50 Matches: 48
Percent Similarity: 33.72% Conservative: 10
Best Local Similarity: 27.91% Mismatches: 73
Query Match: 8.52% Indels: 41
DB: Gaps: 9

US-09-809-545A-2 (1-203) x US-08-343-443B-1 (1-2371)

QY 29 TTTAAGTCTThValleuLeuCySGlnAla---AaNGlNGlGlySerSerMetTySer 47
DB 109 TATGCACAGACACC-----CAGGCATATGGCAACAAAGCTATGGAACCTATGGA 159
QY 48 GTPProSerSerLeuValTyThrSerAlaMetProGlyPheProTyProAlaAlaThr 67
DB 160 CAGCCACGATGTCAGCTATACCCAGGCTCAGACACGACGCAACCTATGGGAGACGCC 219
QY 68 AAlaAlaAlaAlaTyArgGlyAlaHisLeuArgGlyArgGlyArgThrValTyAsnThr 87
DB 220 TATGCACCTCTTAT-----GGACAGCTCCGACGCTGTTACTACT 261
QY 88 PheArgAlaAlaAlaProProProProIleProAlaTyGlyValValTyGlnGlu 107
DB 262 -----CCAACTGCCCCCCAG-----GCATACAGCCAG 288
QY 108 ProValTyGlyAsnLeuLeuGlnGlyTyAlaAlaTyArgTyAlaGlnPro 127
DB 289 CCGTCCAGGGAT-----GGCAGCTGCTTATGATACCAACCACTGCT 333
QY 128 ThrProAlaThrAlaAlaAlaTySerAspSerTyGlyArgValTyAlaAlaAspPro 147
DB 334 ACAGTCACCAACCAAGCC-----TCCATATGACCTCAGTCT 372
QY 148 TyrHisHisThrLeuAlaProAlaProThrTyGlyValGlyAlaMetSerAlaPheAla 167
DB 373 GCATATGGCACTCAGCTGCTTATCCAGCTATGGCAG-----CAG 414
QY 168 ProLeuThrAspAlaAlaTyThrArgSerHisAlaAspAspValGlyLeuValLeuSerSer 187
DB 415 CCAAGCAGCAGCTCAGCTATACAGACCGCAGAGAGGAAACAGCCCACTGAGACTAATCAA 474
QY 188 LeuGlnAlaSerIleTyGlnGlyGlyTyAsnArg 199
DB 475 CCTCAATCTAGC-----ACAGGGGGTTACAAACAG 504

RESULT 9
US-08-437-027-18
/ Sequence 18, Application US/08437027

Patent No. 5670317
GENERAL INFORMATION:
APPLICANT: Landanyi, Marc
APPLICANT: Gerald, William
TITLE OF INVENTION: A DIAGNOSTIC TEST FOR TEST FOR THE DESMOPLASTIC
TITLE OF INVENTION: SMALL ROUND CELL TUMOR
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/437,027
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 46416/JPW/CCA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2412 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
US-08-437-027-18

Alignment Scores:
Pred. No.: 5.89 Length: 2412
Score: 92.50 Matches: 48
Percent Similarity: 33.72% Conservative: 10
Best Local Similarity: 27.91% Mismatches: 73
Query Match: 8.52% Indels: 41
DB: Gaps: 9

US-09-809-545A-2 (1-203) x US-08-437-027-18 (1-2412)

QY 29 TTTAAGTCTThValleuLeuCySGlnAla---AaNGlNGlGlySerSerMetTySer 47
DB 109 TATGCACAGACACC-----CAGGCATATGGCAACAAAGCTATGGAACCTATGGA 159
QY 48 GTPProSerSerLeuValTyThrSerAlaMetProGlyPheProTyProAlaAlaThr 67
DB 160 CAGCCACGATGTCAGCTATACCCAGGCTCAGACACGACGCAACCTATGGGAGACGCC 219
QY 68 AAlaAlaAlaAlaTyArgGlyAlaHisLeuArgGlyArgGlyArgThrValTyAsnThr 87
DB 220 TATGCACCTCTTAT-----GGACAGCTCCGACGCTGTTACTACT 261
QY 88 PheArgAlaAlaAlaProProProProIleProAlaTyGlyValValTyGlnGlu 107
DB 262 -----CCAACTGCCCCCCAG-----GCATACAGCCAG 288
QY 108 ProValTyGlyAsnLeuLeuGlnGlyTyAlaAlaTyArgTyAlaGlnPro 127
DB 289 CCGTCCAGGGAT-----GGCAGCTGCTTATGATACCAACCACTGCT 333
QY 128 ThrProAlaThrAlaAlaAlaTySerAspSerTyGlyArgValTyAlaAlaAspPro 147
DB 334 ACAGTCACCAACCAAGCC-----TCCATATGACCTCAGTCT 372

Qy 148 TyrHisIstHrLeuAlaProAlaProThrTyrglyValGlyAlaMetAsnAlaPheAla 167
Db 373 GCATATGGACTGACCTGCTTATTCAGCTATGGGAG-----CAG 414
Qy 168 ProLeuThrAspAlaLeuThrArgSerHisAlaAspValGlyLeuValLeuSerSer 187
Db 415 CCAGCAGCCACTGACCTTACAAAGACCCAGAGATGGAAACAAAGCCACTGAGACTAGTCAA 474
Qy 188 LeuGlnAlaSerIleTyrglnGlyGlyTyraAsnArg 199
Db 475 CCTCATCTAC-----ACAGGGGGTTACCAACGAG 504
RESULT 10
US-08-258-261B-6
Sequence 6, Application US/08258261B
Patent No. 5639949
GENERAL INFORMATION:
APPLICANT: Schnupp, Thomas
APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Gaffney, Thomas Deane
APPLICANT: Lam, Stephen Ting
APPLICANT: Hammer, Phillip E.
APPLICANT: Uknes, Scott Joseph
TITLE OF INVENTION: Genes for the synthesis of
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/258,261B
FILING DATE: 08-JUN-1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/457,205
FILING DATE: 01-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-258-261B-6
Alignment Scores:
Pred. No.: 169 Length: 28958
Score: 92.50 Matches: 34
Percent Similarity: 41.28% Conservative: 11
Best Local Similarity: 31.19% Mismatches: 52
Query Match: 8.52% Indels: 12
DB: 1 Gaps: 4

US-09-809-545A-2 (1-203) x US-08-258-261B-6 (1-28958)
Qy 47 SerGlyProSerSerLeuValTyThrSerAlaMetProGlyPheProTyProAlaAla 66
Db 10706 AGCGGCTCTCGACGATGGGATCGCTTCCGTCGAGGTACGCCCATCCGTCTCA 10765
Qy 67 ThrAlaAlaAlaAlaTyrr--ArgGlyAlaHis-LeuArgGlyArgGlyArgThrValTy 85
Db 10766 CGCTCGCCCTCGGAGACCTCGAGCGCTCACCGCTGATCCGCTGCTCGGCTCA 10825
Qy 85 rAsmThrPheArgAlaAlaAlaProProProProIleProAlaTyrglyValValTy 105
Db 10826 TTCGACG---AGAGAGAGCCACTCCCGCTGCTCTCTCCGCGGAGGCTCTTA 10882
Qy 105 rGlnGluProValTyrglyAsnLysLeuGlnGlyGlyTyraAlaAlaTyraGlyTyAl 125
Db 10883 CCCGAGGCT-----CGCGCTCGACTGAGAGACTTCTTGGCGCTTACGCTCCCC 10933
Qy 125 agln-----ProThrProAlaThrAlaAlaAlaTyrrSerAspSerTy 139
Db 10934 GCAGAGTCTCCCTCCCGACCTTCCCTTCCAGCGAGAGCGGTCTTGCTGACGCTTCA 10993
Qy 139 rGlyArgValTyraAlaAspPro 147
Db 10994 CGGACGAACGCTTCCGAGTCCGCT 11018
RESULT 11
US-08-456-837-6
Sequence 6, Application US/08456837
Patent No. 5643774
GENERAL INFORMATION:
APPLICANT: Schnupp, Thomas
APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Gaffney, Thomas Deane
APPLICANT: Lam, Stephen Ting
APPLICANT: Hammer, Phillip E.
APPLICANT: Uknes, Scott Joseph
TITLE OF INVENTION: Genes for the synthesis of
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,837
FILING DATE: 01-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/457,205
FILING DATE: 01-JUN-1995
APPLICATION NUMBER: 08/258,261
FILING DATE: 08-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:

```

? SEQUENCE CHARACTERISTICS:
? LENGTH: 28958 base pairs
? TYPE: nucleic acid
? STANDARDS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHEetical: NO
? ANTI-SENSE: NO
US-08-456-837-6

```

Pred. No.:	169	Length:	28956
Score:	92.50	Matches:	34
Percent Similarity:	41.28%	Conservative:	11
Best Local Similarity:	31.19%	Mismatches:	52
Query Match:	8.52%	Indels:	12
DB:	1	Gaps:	4

US-09-809-545A-2 (1-203) X US-08-456-837-6 (1-28958)

QY	47	SEGLYPRSESR	euVal1YrThrSerAlaMetProGlyPheProTyrrProAlaala	66
Db	10706	ACCGGCTCTCCAGC	ATGGGCATCGTTTCTCCGTGCAGGCTACGCCCATCCCGTCCTCA	107655
QY	67	ThrAlaAlaAlaAla	lery---ArgGlyAlaHis-LeuArgGlyArgGlyArgThrValTy	85
		:::		
Db	10766	CGCTCGCCCTCCGCGAG	ACTGTGCAGACGCTCACCCGTGATCCCGTCGTCGTCGCTCA	108235
QY	85	rAsnThrPheArgAlaAlaAla	leProProProProIleProAlaTYrGlyValValTy	105
		:::		
Db	10826	TTTCAGC---AGAGAGAG	CCACCTCGCCGCTGCTCTCTTCCTGGCGGAGCTCTCTA	108825
QY	105	rGlnGlnPProValTYrGly	AsnLysLeuLeuGlnGlyGlyTYrAlaAlaAlaTYrArgTYrAl	125
Db	10883	CCCGAGGCT-----	CGCGCTCGACTGGAGGACTTCTTCGGCGCCCTACGCTCCCC	109335
QY	125	Asn-----	ProThrProAlaThrAlaAlaAlaTYrSerAspSerTy	139
Db	10934	GGAGAGTTCCTCCCTCC	CCACCTACCCCTTCACAGGAGAGCGGTTTGCGTCGACGTCTCCA	109995
QY	139	rGlyArgValTYrAlaAla	AspPro	147
		:::		
Db	10994	CGAGCGAACGCTTCG	ACGCTGCTT	11018

RESULT 12
 US-08-457-342-6
 : Sequence 6, Application US/08457342
 : Patent No. 5662898
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Schupp, Thomas
 :
 : APPLICANT: Ligon, James M.
 :
 : APPLICANT: Beck, James Joseph
 :
 : APPLICANT: Hill, Dwight Steven
 :
 : APPLICANT: Ryals, John Andrew
 :
 : APPLICANT: Gaffney, Thomas Deane
 :
 : APPLICANT: Lam, Stephen Ting
 :
 : APPLICANT: Hammer, Phillip E.
 :
 : APPLICANT: Uknes, Scott Joseph
 :
 : TITLE OF INVENTION: Genes for the synthesis of
 :
 : TITLE OF INVENTION: antipathogenic substances
 :
 : NUMBER OF SEQUENCES: 22
 :
 : CORRESPONDENCE ADDRESS:
 :

```

1 CURRENT APPLICATION DATA:
2 APPLICATION NUMBER: US/08/457,342
3 FILING DATE: 01-JUN-1995
4 CLASSIFICATION: 424
5 PRIOR APPLICATION DATA:
6 APPLICATION NUMBER: US 08/457,205
7 FILING DATE: 01-JUN-1995
8 APPLICATION NUMBER: 08/258,261
9 FILING DATE: 08-Jun-1994
10 ATTORNEY/AGENT INFORMATION:
11 NAME: Elmer James Scott
12 REGISTRATION NUMBER: 36,129
13 REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
14 TELECOMMUNICATION INFORMATION:
15 TELEPHONE: 919-541-8614
16 TELEFAX: 919-541-8689
17 INFORMATION FOR SEQ. ID NO: 6:
18 SEQUENCE CHARACTERISTICS:
19 LENGTH: 28958 base pairs
20 TYPE: nucleic acid
21 STRANDEDNESS: single
22 TOPOLOGY: linear
23 MOLECULE TYPE: DNA (genomic)
24 HYPOTHEetical: NO
25 ANTI-SENSE: NO
26 US-08-457-342-6

```

Pred. No.:	169	Length:	2895
Score:	92.50	Matches:	34
Percent Similarity:	41.48%	Conservative:	11
Best Local Similarity:	31.19%	Mismatches:	52
Query Match:	8.52%	Indels:	12
DB:	1	Gaps:	4

US-09-809-545A-2 (1-203) X US-08-457-342-6 (1-28958)

[illegible]

RESULT 13
US-08-457-645A-6
Sequence 6, Application US/08457645A
Patent NO. 5679560
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Liqon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Gaffney, Thomas Deane
APPLICANT: Lam, Stephen Ting

```

APPLICANT: Hammer, Phillip E.
APPLICANT: Uknes, Scott Joseph
TITLE OF INVENTION: Genes for the synthesis of
TITLE OF INVENTION: antipathogenic substances
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/457,646A
FILING DATE: 01-JUN-1995
CLASSIFICATION: 530
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/457,205
FILING DATE: 01-JUN-1995
APPLICATION NUMBER: 08/258,261
FILING DATE: 08-Jun-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-457-646A-6

Alignment Scores:
Pred. No.: 169 Length: 28958
Score: 92.50 Matches: 34
Percent Similarity: 41.28% Conservative: 11
Best Local Similarity: 31.19% Mismatches: 52
Query Match: 8.52% Indels: 12
DB: 1 Gaps: 4

US-09-809-545A-2 (1-203) x US-08-457-646A-6 (1-28958)
Qy 47 SerGlyProSerSerLeuValTyrrThSerAlaMetProGlyPheProTyrrProAlaAla 66
Db 10706 AGCGGCTCTCGACGATGGGATCGCTTCTCCGTGAGGTGACGCCCATCCCGGCTCA 10765
Qy 67 ThrAlaAlaAlaAlaTyrr--ArgGlyAlaHis-LeuArgGlyArgGlyArgThValTy 85
Db 10766 CGCTCGCGCTCGGAGACCTGCGAGCGCTCACCCGCTGATCCGCTGCTCGGCTCA 10825
Qy 85 rAsnThrPheArgAlaAlaAlaProProProProProProProProProProProProPro 105
Db 10826 TTCGACG--AGAGAGAGGACCTCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10882
Qy 105 rGlnGluProValTyrrGlyAsnLysLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 125
Db 10883 CCGGAGGCGCT-----CGGCTCGACCTGAGAGACTTCTCCGCGCTCTACGCTCCCC 10933
Qy 125 agIn-----ProThrProAlaThrAlaAlaAlaTyrrSerAspSerTy 139
Db 10934 GCAAGGTCTCCCTCCCACTACCTTCCAGCGAGAGCGGTCTGCTCGACGTCTCCA 10993

```

```

Qy 139 rGlyArgValTyrrAlaAlaAspPro 147
Db 10994 CGAGCAACGCTTCCGACGTGCGCT 11018

RESULT 14
US-08-458-076A-6
Sequence 6, Application US/08458076A
Patent No. 5698425
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James M.
APPLICANT: Beck, James Joseph
APPLICANT: Hill, Dwight Steven
APPLICANT: Ryals, John Andrew
APPLICANT: Gaffney, Thomas Deane
APPLICANT: Lam, Stephen Ting
APPLICANT: Hammer, Phillip E.
APPLICANT: Uknes, Scott Joseph
TITLE OF INVENTION: Genes for the synthesis of
TITLE OF INVENTION: antipathogenic substances
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,076A
FILING DATE: 01-JUN-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/457,205
FILING DATE: 01-JUN-1995
APPLICATION NUMBER: 08/258,261
FILING DATE: 08-Jun-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-458-076A-6

Alignment Scores:
Pred. No.: 169 Length: 28958
Score: 92.50 Matches: 34
Percent Similarity: 41.28% Conservative: 11
Best Local Similarity: 31.19% Mismatches: 52
Query Match: 8.52% Indels: 12
DB: 1 Gaps: 4

US-09-809-545A-2 (1-203) x US-08-458-076A-6 (1-28958)
Qy 47 SerGlyProSerSerLeuValTyrrThSerAlaMetProGlyPheProTyrrProAlaAla 66

```

```

Db 10706 AGCGCTCTCGACGATGGGCAATCGCTTCCGCTGAGGTACAGCCCAATCCCGTCTCA 10765
Qy 67 ThrAlaAlaAlaAlaTyr---ArgGlyAlaHis-LeuArgGlyArgGlyArgThxValTy 85
Db 10766 CGCTGCGCTCCGCGAGACCTGGAGCGCTCAACCGCTGCATCCCGTCGTCGCTCCA 10825
Qy 85 rAsnThrPheArgAlaAlaAlaProProProProlleProAlaTyrglyValValTy 105
Db 10826 TTCGACG---AGAGAGAGGCCACCTGCGCGCTGCTCTCTCTGCGGCGAGCTCTCA 10882
Qy 105 rGlnGluProValTyrglyAsnlyLeuLeuGlnGlyTyrglyAlaAlaTyrglyTyra 125
Db 10883 CCCGAGGCGCT-----CGGCTCGACTCGAGAGACTTCTTCCGCGCTTACGCTCCCC 10933
Qy 125 agln-----ProThrProAlaThrAlaAlaAlaTyrglySerAspSerTy 139
Db 10934 GGAAGTCTCCTCCCTCCACCTACCTTCCAGAGAGAGGGGTTCTGCTCGACGCTCCA 10993
Qy 139 rGlyArgValTyrglyAlaAlaAspPro 147
Db 10994 CGAGCGAAGCTTCCGAGCTCGCT 11018

RESULT 15
US-08-764-233A-4
; Sequence 4, Application US/08764233A
; Patent No. 5716849
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Schupp, Thomas
; APPLICANT: Beck, James J.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Neff, Snezana
; APPLICANT: Ryals, John A.
; TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,233A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
; FILING DATE: 09-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: 1506/CIP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8587
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Sorangium cellulosum

```

```

; IMMEDIATE SOURCE:
; CLONE: p98/1
; US-08-764-233A-4

Alignment Scores:
Pred. No.: 169 Length: 28958
Score: 92.50 Matches: 34
Percent Similarity: 41.28% Conservative: 11
Best Local Similarity: 31.19% Mismatches: 52
Query Match: 8.52% Indels: 12
DB: 1 Gaps: 4

US-09-809-545a-2 (1-203) x US-08-764-233A-4 (1-28958)

Qy 47 SerGlyProSerSerLeuValTyrglySerAlaMetProGlyPheProTyrglyProAla 66
Db 10706 AGCGCTCTCGACGATGGGCAATCGCTTCCGCTGAGGTACAGCCCAATCCCGTCTCA 10765
Qy 67 ThrAlaAlaAlaAlaTyr---ArgGlyAlaHis-LeuArgGlyArgGlyArgThxValTy 85
Db 10766 CGCTGCGCTCCGCGAGACCTGGAGCGCTCAACCGCTGCATCCCGTCGTCGCTCCA 10825
Qy 85 rAsnThrPheArgAlaAlaAlaProProProProlleProAlaTyrglyValValTy 105
Db 10826 TTCGACG---AGAGAGAGGCCACCTGCGCGCTGCTCTCTCTGCGGCGAGCTCTCA 10882
Qy 105 rGlnGluProValTyrglyAsnlyLeuLeuGlnGlyTyrglyAlaAlaTyrglyTyra 125
Db 10883 CCCGAGGCGCT-----CGGCTCGACTCGAGAGACTTCTTCCGCGCTTACGCTCCCC 10933
Qy 125 agln-----ProThrProAlaThrAlaAlaAlaTyrglySerAspSerTy 139
Db 10934 GGAAGTCTCCTCCCTCCACCTACCTTCCAGAGAGAGGGGTTCTGCTCGACGCTCCA 10993
Qy 139 rGlyArgValTyrglyAlaAlaAspPro 147
Db 10994 CGAGCGAAGCTTCCGAGCTCGCT 11018

```

Search completed: March 16, 2003, 04:02:27
Job time : 114 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: March 16, 2003, 02:35:26 ; Search time 105 Seconds

(without alignments)
1356,939 Million cell updates/sec

Title: US-09-809-545A-2

Perfect score: 1086
Sequence: 1 MTMKKAVNPYTNQMKLNPPV.....VLSSLSQASIVQGYNRPAPY 203

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 501302 seqs, 350932545 residues

Total number of hits satisfying chosen parameters: 1002604

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=cgnt1/USPFO.spool/US09809545/runat_07032003.153854.11508/asp_query.fasta_1.391
-DB=Published Applications NA -QFMT=fastcap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSTZ=500 -MINLEN=0
-MALEN=200000000 -USER=US09809545 @CGN 1 1 80 @runat_07032003.153854.11508
-NCPU=6 -ICPU=3 -NO_XLIPX -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

1:	/cgnt2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2:	/cgnt2_6/ptodata/2/pubpna/PTC1_NEW_PUB.seq:*
3:	/cgnt2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4:	/cgnt2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5:	/cgnt2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6:	/cgnt2_6/ptodata/2/pubpna/PTCUTS_PUBCOMB.seq:*
7:	/cgnt2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8:	/cgnt2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9:	/cgnt2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10:	/cgnt2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11:	/cgnt2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12:	/cgnt2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13:	/cgnt2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14:	/cgnt2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1086	100.0	1340	10	US-09-809-545A-1
2	955	87.9	2372	10	US-09-794-591-1
3	108.5	10.0	6930	9	US-10-098-841-226
4	107	9.9	895	10	US-09-770-445-481

5	98.5	9.1	821	9	US-09-996-634-62	Sequence 62, Appl
6	97.5	9.0	2658	10	US-09-815-242-4035	Sequence 4035, Ap
7	97.5	9.0	3285	9	US-09-712-363-143	Sequence 143, App
8	96	8.8	3231	8	US-08-754-311B-1	Sequence 1, Appl1
9	94	8.7	657	10	US-09-974-300-1655	Sequence 1655, Ap
10	92.5	8.5	2273	10	US-09-822-830A-410	Sequence 410, App
11	92.5	8.5	2350	10	US-09-880-107-3769	Sequence 3769, Ap
12	92	8.5	53522	9	US-09-904-968A-1	Sequence 1, Appl1
13	91.5	8.4	969	9	US-10-125-635A-452	Sequence 452, App
14	90.5	8.3	1743	10	US-09-887-576-788	Sequence 788, App
15	90	8.3	1782	9	US-09-938-842A-1143	Sequence 1143, Ap
16	90	8.3	2303	10	US-09-880-107-3849	Sequence 1849, Ap
17	90	8.3	62944	10	US-09-954-456-2257	Sequence 2257, Ap
18	89.5	8.2	974	10	US-09-864-761-9913	Sequence 9913, App
19	89.5	8.2	32249	10	US-09-764-878-202	Sequence 202, App
20	89	8.2	1167	9	US-09-938-842A-2608	Sequence 8001, Ap
21	88.5	8.1	540	10	US-09-864-761-8001	Sequence 453, App
22	88.5	8.1	1410	9	US-10-125-635A-453	Sequence 201, App
23	88.5	8.1	32190	10	US-09-764-878-201	Sequence 200, App
24	88.5	8.1	32193	10	US-09-764-878-200	Sequence 3, Appl1
25	87.5	8.1	1917	9	US-10-038-010-3	Sequence 186, Ap
26	87.5	8.1	2298	9	US-09-954-531-1386	Sequence 551, App
27	87	8.0	2610	10	US-09-925-301-551	Sequence 2, Appl1
28	87	8.0	12606	10	US-09-957-974-2	Sequence 1372, Ap
29	86.5	8.0	948	10	US-09-917-800A-244	Sequence 244, App
30	86.5	8.0	14136	9	US-09-964-878A-1244	Sequence 3, Appl1
31	86.5	8.0	80959	9	US-09-858-546-3	Sequence 19, Appl1
32	86	7.9	1801	10	US-09-796-858-19	Sequence 5, Appl1
33	86	7.9	2046	10	US-09-782-906-7	Sequence 389, App
34	86	7.9	4020	9	US-09-796-679-5	Sequence 107, App
35	85	7.8	1263	9	US-10-125-635A-389	Sequence 183, App
36	85	7.8	1497	10	US-09-060-854B-1	Sequence 580, App
37	85	7.8	1707	9	US-10-125-635A-390	Sequence 289, App
38	85	7.8	24533	9	US-09-764-868-1349	Sequence 1171, Ap
39	84.5	7.8	1208	10	US-09-764-846-93	Sequence 93, Appl
40	84.5	7.8	1520	9	US-09-924-340-107	Sequence 107, App
41	84.5	7.8	1520	9	US-09-992-600A-107	Sequence 183, App
42	84.5	7.8	1528	9	US-09-746-783-183	Sequence 580, App
43	84.5	7.8	3067	10	US-09-925-301-580	Sequence 289, App
44	84.5	7.8	3118	10	US-09-864-864-289	Sequence 1171, Ap
45	84.5	7.8	3314	10	US-09-954-456-1171	Sequence 1171, Ap

ALIGNMENTS

RESULT 1
US-09-809-545A-1
; Sequence 1, Application US/09809545A
; Patent NO. US20020110804A1
; GENERAL INFORMATION:
; APPLICANT: Stanton, Lawrence W.
; TITLE OF INVENTION: SECRETED FACTORS
; FILE REFERENCE: SCIOS.017A
; CURRENT APPLICATION NUMBER: US/09/809,545A
; CURRENT FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PasteSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1340
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-09-809-545A-1

Alignment Scores:
Pred. No.: 1,48e-98 Length: 1340
Score: 1086.00 Matches: 203
Percent Similarity: 100.00% Conserves: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0
US-09-809-545A-2 (1-203) x US-09-809-545A-1 (1-1340)

```
QY 1 MetThrAsnLysLeuAlaValAsnProTyrThrAsnGlyTTrpLysLeuAsnProValVal 20
DB 535 ATGACTAATATAAAAGGCGCGTGAACCCCTACACCAATGGCTGGAAATTAAATCAGTTGTG 594
QY 21 G1YAlaValATySerProAspPheTyrAlaGlyThrValLeuLeuCyGlnAlaAsnGln 40
DB 595 GCGCGGCTGTACAGCCCGCATCTTATGACAGGACGGTGTGTGTGCCAGGCGCAACCGAG 654
QY 41 GIUGLysSerSerMetTyrSerGlyProSerSerLeuValATyThrSerAlaMetProGly 60
DB 655 GAGGATCTTCCATGATACAGTGGCCCCAGTTCCTGTATATCTTCTGCAATGGCTGGC 714
QY 61 PheProTyrProAlaAlaThrAlaAlaAlaAlaATyArgGlyValHisLeuArgGlyArg 80
DB 715 TTTCATATCCGCGCGCCACTGCTGCAGCTGCATACCGAGGGGCTCACCTTGCAGGCCGT 774
QY 81 G1YArgThrValATyArgThrPheAlaGlnAlaAlaProProProProTyrProAlaTyr 100
DB 775 GGTCCGACCGGTACAAACCTTCAGAGCTGCGCGCCCGCCCAATCCCGGCTAT 834
QY 101 G1YGIValValATyGlnGluProValATyGlyAsnLysLeuGlnGlyTyrAla 120
DB 835 GCGGAGTGTGTATCAAGAGCCAGTGTATGCAATTAATGCTACAGGGGTGTTACGCT 894
QY 121 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaATySerAspSerTyrGly 140
DB 895 GCATACCGCTACGCCCGACCCCTGCGCACCTGCTGCCTACAGTACAGTACGGA 954
QY 141 ArgValATyAlaAlaAspProTyrHisLeuThrLeuAlaProAlaProThrTyrGlyVal 160
DB 955 CAGATTATGCTGCGCCAGCCCTTACCAACACCTTGTCCACGCCCGCCCTACAGCGCTT 1014
QY 161 G1YAlaMetAsnAlaPheAlaProLeuThrAspAlaLeuThrArgSerHisAlaAspAsp 180
DB 1015 GGTGCATATAAGCTTTTGGCGCTTGACCGATGCCAAGATGGAGCCATGCTGAT 1074
QY 181 ValGlyLeuValLeuSerLeuGlnAlaSerIleTyrGlnGlyGlyTyrAsnArgPhe 200
DB 1075 GTGGGCTGCTCTTCTTTCATGACAGGCTAGATATACCAAGGGGATACAAACGTTT 1134
QY 201 AlaProTyr 203
DB 1135 GCTCCATAT 1143

RESULT 2
US-09-794-591-1
; Sequence 1, Application US/09794591
; Patent No. US20010018198A1
; GENERAL INFORMATION:
; APPLICANT: Pulset, Stefan M.
; APPLICANT: Shibata, Hiroki
; TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
; TITLE OF INVENTION: Products Related Thereto and Methods of Using Same
; FILE REFERENCE: CE 3093
; CURRENT APPLICATION NUMBER: US/09/794,591
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/145,391
; PRIOR FILING DATE: 1998-09-01
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2372
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (987) .. (1979)
US-09-794-591-1

Alignment Scores: 2,63e-85 Length: 2372
Pred. No.: 955.00 Matches: 169
```

```
Percent Similarity: 86.94%
Best Local Similarity: 85.14%
Query Match: 87.94%
DB: 10 Gaps: 2

US-09-809-545A-2 (1-203) x US-09-794-591-1 (1-2372)

QY 1 MetThrAsnLysLeuAlaValAsnProTyrThrAsnGlyTTrpLysLeuAsnProValVal 20
DB 1572 ATGACAAATATAAAAGCGGTCAACCTTATACAAATGGCTGGAAATTGAATCAGTTGTG 1631
QY 21 G1YAlaValATySerProAspPheTyrAlaGlyThrValLeuLeuCyGlnAlaAsnGln 40
DB 1632 GGTGAGTCTTACAGTCCGGAATTCTATGACGAGCGTCTGTGTGTGCCAGGCGCAACCG 1691
QY 41 GIUGLysSerSerMetTyrSerGlyProSerSerLeuValATyThrSerAlaMetProGly 60
DB 1692 GAGGATCTTCCATGATACAGTGGCCCCAGTTCCTGTATATCTTCTGCAATGGCTGAGC 1751
QY 61 PheProTyrProAlaAlaThrAlaAlaAlaAlaATyArgGlyValHisLeuArgGlyArg 80
DB 1752 TTCCGATCCAGACGACCGCCCGCGGCTACCGAGGGGCGCACCTGCGAGGCCGC 1811
QY 81 G1YArgThrValATyArgThrPheAlaGlnAlaAlaProProProProTyrProAlaTyr 100
DB 1812 GGTCCGACCGGTACAAACCTTCAGAGCGCGCGCGCCCGCCCAATCCCGGCTAC 1871
QY 101 G1YGIValValATyGlnGluProValATyGlyAsnLysLeuGlnGlyTyrAla 120
DB 1872 GCGGCTGTTTACCCAGAGATGATTTTATGCTGACAGC---ATTATGAGTGTATGC 1928
QY 120 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaAlaATySerAspSer---- 138
DB 1929 TCATATCCGCTACGCCCGACGCTTACCCCTGCGCACCTGCTGCCTACAGTACAG-AAAATC 1987
QY 139 -----TyrGlyArgVa 142
DB 1988 AGTTCGCTCTGCTGACAGATGAATTTCTTGAACCTCTGCAGTTACGAGCAT 2047
QY 142 TTYrAlaAlaAspProTyrHisLeuThrLeuAlaProAlaProThrTyrGlyValAla 162
DB 2048 TTATGCTGCGGACCCCTACACCAAGCACTTCTCAGGCCCGCCACCTACGCGTTGCTGC 2107
QY 162 aMetAsnAlaPheAlaProLeuThrAspAlaLeuThrArgSerHisAlaAspAspValG 182
DB 2108 CATGAATGCTTTTGCACTTGTACATGATGCCAAGACTAGAGCCATGCTGATGATGTGG 2167
QY 182 YLeuValLeuSerSerLeuGlnAlaSerIleTyrGlnGlyGlyTyrAsnArgPheAlaPr 202
DB 2168 TCTCGTCTTCTTTCATGACAGGCTAGATATACCGAGGGGATACAAACGTTTGTCTCC 2227
QY 202 cTyr 203
DB 2228 ATAC 2231

RESULT 3
US-10-098-841-226
; Sequence 226, Application US/10098841
; Publication No. US20020197679A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Xu, Chongjun
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunding
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhao, Qing A.
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wenman, Tom
```

```

; APPLICANT: Zhang, Jie
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Dimanac, Radoje T.
; TITLE OF INVENTION: No. US20020197679A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2
; CURRENT APPLICATION NUMBER: US/10/098, 841
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: 09/598, 042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/552, 317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488, 725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 331
; SOFTWARE: pt_fl_genes Version 1.0
; SEQ ID NO 226
; LENGTH: 6930
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (234)..(968)
; US-10-098-841-226

Alignment Scores:
Pred. No.: 0.405 Length: 6930
Score: 108.50 Matches: 56
Percent Similarity: 35.85% Conservative: 20
Best Local Similarity: 26.42% Mismatches: 67
Query Match: 9.99% Indels: 12
DB: Gaps: 69

US-09-809-545A-2 (1-203) x US-10-098-841-226 (1-6930)
Qy 6 AlaValaAsnProTyThrAsnGlyTrpIysLeuAsnProValaValGlyAlaValTyrSer 25
Db 336 GCCCTACATATCC-----AGCTGTACCCCAATATGATCCAGTTATCT 380
Qy 26 ProAspPhe-----TyrAlaGlyThrValLeuLeuCyGlnAlaAsn 39
Db 381 CCAGAGTTTCAGTTCTCTGATTCAGCTTATGCA---ACTGTGATGATGAAGGCGCTGG 437
Qy 40 GlnGluGlySerSer-----MetTyrSerGly 48
Db 438 CCACAGAACTGCTTCTCTGTCGACGTGAAGCACTTCCACCTCCAGTGCACCGG 497
Qy 49 ProSerSerLeuValTyrThrSerAlaMetProGlyPheProTyrProAlaAlaThrAla 68
Db 498 ACCGAGAACCAACTTACCAAGCATCTCTGCGCTTCAGATAT-----ACTGCG 548
Qy 69 AlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArgGlyThrValTyrAsnThrPhe 88
Db 549 GGGACACCATACAG-----GTCCACCGGAGCCAG 578
Qy 89 ArgAlaAlaAlaProProProProleProAlaTyrGlyValValTyrGlnGlnPro 108
Db 579 AGTAACACTGCTCCACCCCTCTACTCTCCATCACCAC-----CCTATACAGCGGC 632
Qy 109 ValTyr-----GlyAsnLysLeuLeuGlnGlyGlyTyrAla 120
Db 633 ATGTATCAATCAGAAAGTCTTACCCCGAGATCTGTATGCGACGAGACCTAC--- 689
Qy 121 AlaTyrArgTyrAlaGlyProThrProAlaThrAlaAlaTyrSerAspSerTyrGly 140
Db 690 -----TACACACAGCGC----- 701
Qy 141 ArgValTyrAlaAlaAspProTyr-----HisHisThrLeuAlaProAlaProThrTyr 158
Db 702 ---GCTATGCTGCGCCAGCTCATGTCATCCACACACAGCGTGTCCAGCCCAAC--- 755
Qy 159 GlyValGlyAlaMetAsnAlaPheAlaProLeuThrAspAlaLysThrArgSerHisAla 178

```

```

Db 756 -----ACATTCCTCTGCTATCTACCCAGACCTGTGCGCCCGGAGAGACCAAGCT 809
Qy 179 AspAspValGlyLeuValLeuSerSerLeuGlnAla 190
Db 810 GTGGCATGGGCGATGTGTGGAGGACCAACCATGGCA 845

RESULT 4
US-09-770-445-481/c
; Sequence 481, Application US/09770445
; Patent No. US20020023281A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Jong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Moesner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hubban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2023US (PARA-012PRV)
; CURRENT APPLICATION NUMBER: US/09/770, 445
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178, 472
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 481
; LENGTH: 895
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-770-445-481

Alignment Scores:
Pred. No.: 0.0436 Length: 895
Score: 107.00 Matches: 52
Percent Similarity: 36.60% Conservative: 19
Best Local Similarity: 26.80% Mismatches: 63
Query Match: 9.85% Indels: 60
DB: Gaps: 11

US-09-809-545A-2 (1-203) x US-09-770-445-481 (1-895)
Qy 42 GlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGlyPhe 61
Db 869 GGTATATGCGCTTACCAAGTCCA-----ACACGACGACGACCGTCATAT 825
Qy 62 ---ProTyrProAlaAla-----ThrAlaAlaAlaTyrArgGlyAla 75
Db 824 GATTCAACACGACGACGACATGATGTTCAACTCCAGCAGACCATCTGATGTTCA 765
Qy 76 HisLeuArgGlyArg-----GlyArgThrValTyr 85
Db 764 AACATGGCTCAACACAGCAATATGTTATCATCACTGAGCGCTTACTCAGCAACTTAT 705
Qy 86 AsnThrPheArgAlaAlaAlaPro-----ProProProle 97
Db 704 CCTTCATATAGCTGTGACGACCATCTGATGTTATATGTAACACACACGACAGTT 645
Qy 98 ---ProAlaTyr-----GlyGlyValValTyrGlnGlnPro 108

```

```

Db 644 GCCCAGCTTATGACGACGAGCTGCTCAGCCAGCTTCTGCTGTCGACAACTTCAGGT 585
Qy 109 VALTYGlyAenlyeuleuenginglylyTyraalaTyraTyraTyraTyraTyraTyra 128
Db 584 GGGTACGGGCAAGTACCTCCACCGGTGGCTATAGTTCGTATCCCTCCACAGCGCGGT 525
Qy 129 ProAlaThrAlaAlaAlaTySerAsp---SerTyGlyArgValTyraAlaAlaAspPro 147
Db 524 TATGTAATACCCCGCTCAAGCAATGAACTATGGA-----TACATTGGCTCTCAG 471
Qy 148 TyrHisLeuThrLeuAlaProAlaProThrTyGlyValGlyAlaMetAsnAlaPheAla 167
Db 470 TAT-----CCTAGCTATGGAGGTGGAACCATCGCATATGCT 432
Qy 168 ProLeuThrAspAlaLeuThrArgSerHisAlaAspAspValGlyLeuValLeuSerSer 187
Db 431 GACCTACCTGGCAACCGCTTATTC-----TACATTGGCTCTCAG 405
Qy 188 LeuGlnAlaSerIleTyGlyGlyTyraAsnArgPheAla 201
Db 404 ---CAGACTGACACTCTCCAGCGCGCTATGACAAATCAGCA 366

```

RESULT 5

```

US-09-996-634-62
; Sequence 62, Application US/09996634
; Patent No. US20020172684A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; FILE REFERENCE: 61260
; CURRENT APPLICATION NUMBER: US/09/996,634
; PRIOR FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 821
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(821)
; OTHER INFORMATION: n is a, c, g, or t/u.
US-09-996-634-62

```

Alignment Scores:

```

Pred. No.: 0.27 Length: 821
Score: 96.50 Matches: 59
Percent Similarity: 28.46% Conservative: 15
Best Local Similarity: 22.69% Mismatches: 95
Query Match: 9.07% Indels: 91
DB: Gaps: 9

```

US-09-809-545A-2 (1-203) x US-09-996-634-62 (1-821)

```

Qy 1 MetThrAsnLysLysAlaValaAspProTyThrAsnGlyTyRlyLeuAsnProValVal 20
Db 70 ATGGCGGACGAGCTGCTGGCCGAGCCGACCCCAATGCCGCACTGCTGCAACCGGTTCCG 129
Qy 21 GlyAlaValaTySerProAspPheTyraGlyThrValLeuLeuGlyGlnAlaAsnGln 40
Db 130 GCGCAGCGCTTCGACCGACGAGACCTGGCGGTATCACTCCGTCGCTTCAAAACCG 189
Qy 41 GluGlySerMetTySerGlyProSer-----SerLeuValTyR 54

```

```

Db 190 AGGCGCTGGCGCAGAGACTCAAGTCCGACCCGGTGCTCCAAACCCGGGCTGGTCAAC 249
Qy 55 ThrSerAlaMetProGlyPheProTyProAlaAlaThrAlaAlaAlaTyraGly 74
Db 250 TCCGATGGCTGCCCAACAAACCAACCGCGCATCACCAGCTCCGCGGGGACCCCGGA 309
Qy 74 ----- 74
Db 310 GGAAGGCGCCGAGNTCGGATNCAAGGGTTGCMACGGCGCGCTGCCGTTCNGATTGGA 369
Qy 75 -----AlaHisLeuArgGlyArg-----GlyArgThrValTyR 85
Db 370 CCCGACAGCTACCCCGGTATGGGACGCTACGGGAGAAACAACCTGGCCGACCGCCAC 429
Qy 86 AsnThrPheArgAlaAlaAlaProProProIleProAlaTyGlyValValValTyR 105
Db 430 CTCGGCTGTGACAGTTACCGCCCGCGAGCCGAGACCGGCGCGTGTGT----- 480
Qy 106 GlnGluProValTyGlyAsnLys-----LeuLeuGlnGlyGly----- 118
Db 481 -----GGTTCCGGCGCGCGCATCTGTCTTCAAGAGAGACGGCATTTGATCTTA 534
Qy 119 -----TyraAlaAlaTyraGlyTyraGlyProAlaThr 131
Db 535 CGGCACTCCCTGAAACTGCACTGAGGGGCTCACCGGCCGAGCGCGCATTCACGACT 594
Qy 132 AlaAlaAlaTySerAspSerTyR----- 139
Db 595 GGGGAGGATTTCCGATGACATCGAGACCGGACCGCGGTGCGCAATCTCGGTTCC 654
Qy 140 -----GlyArgValTyraAlaAlaAspProTyHisHis----- 150
Db 655 GCTGGCTGGGCGCGCGAGCGGACGCTGCGG-----CATGTGCG 699
Qy 151 -----ThrLeuAlaProAlaProThrTyGly 159
Db 700 CATGACCCGAACTGAGCGCTGAGCAATGTTCCCTTACCCCGCGGTTCCGCT 759
Qy 160 ValGlyAlaMetAsnAlaPheAlaProLeuThrAspAlaLeuThrArgSerHisAlaAsp 179
Db 760 GCTGGAATCTGACAGCGGTGATCGGTCAGCAGACCGGTTGATGATGATGATGATGATGAT 819

```

RESULT 6

```

US-09-815-242-4035/c
; Sequence 4035, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308

```

```

; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4035
; LENGTH: 2658
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-815-242-4035

Alignment Scores:
Pred. No.: 148 Length: 2658
Score: 97.50 Matches: 60
Percent Similarity: 36.19% Conservative: 16
Best Local Similarity: 28.57% Mismatches: 74
Query Match: 8.98% Indels: 61
DB: 10 Gaps: 10

US-09-809-545A-2 (1-203) x US-09-815-242-4035 (1-2658)

Qy 14 TrpLysLeuAnProVala1G1yAlaValTyrSerProAspPheTyrAlaG1yThrVal 33
Db 2263 TGGCAGCATGCTGCTGACCCGGTACGGCGCCAGCACCG-----CGGCGAGCGGCTTCA 2210

Qy 34 LeuLeuCysGln-----Ala 38
Db 2209 GCGGCTGCTCCACGATGTCGCCCGCCACCCAGTCCGCGCCAGCTTCCAGCCACCGT 2150

Qy 39 AsnGlnGluGlySer-----SerMetTyrSerGlyProSerSerLeuValTyr 54
Db 2149 GCGCGAGCGGGGTGCTGTCGACGACGATTCGTGATATAGCGGCTCCAGTCTCGGCTCGT 2090

Qy 55 ThrSerAlaMetPro-----GlyPheProTyrProAla--- 65
Db 2089 CGCGGCTCTTTCACGACGATTCGCGAGATCCGCGGGGATGTCCTCACCCAGCGCCA 2030

Qy 66 -----AlaThrAlaAlaAlaTyrArgGlyAlaHis 76
Db 2029 GCAGGCTGCTGATGAGAGCCGCGCATGCGCTGACGCGGCTGCGCGAGTGTGGAACCG 1970

Qy 77 LeuArgGlyArgGlyArgThrValTyrAsnThrPheArgAlaAlaAlaProProPro--- 95
Db 1969 AGGCGAGCAAGGCGCTGCGCGAGTTCCTCGGCTCGCGCGAGCGCGGCTCTCAAGT 1910

Qy 96 -----ProIleProAlaTyrGlyValValTyrGlnProValTyr 110
Db 1909 CCTCGGCGAGCGAGCCCGCGCCGACGGCTGG-----CCAGAGCTGCGCGA 1862

Qy 111 GlyAsnLysLeuLeuGlnGly-----GlyTyrAlaAlaTyrArgTyrAlaGlnProThr 128
Db 1861 GGGCGCGAGAGCGCGCGCGCTCGGGCTGACGCGCTGCGCGCTGCGAGCGGC-CCGACA 1803

Qy 129 ProAlaThrAlaAlaAlaTyrSerAspSerTyrGlyArgValTyrAlaAlaAspProTyr 148
Db 1802 CCGAGCATGCGACGAGGCTGCTCT-----TCCGCGACAGCGCGC 1764

Qy 149 HisHisThrLeuAlaProAlaProAlaProThrTyrGlyVal-----GlyAlaMetAsnAlaPhe 166
Db 1763 CACCAACCAACGAGCTCTCTCGGCGACGCGTGCACGACGAGCGCGCGCTGCTGCG 1704

Qy 167 AlaProLeuThrAspAlaLysThrArgSer 176
Db 1703 GCCCAGGCCAGTCCGCGGCGACCGCTTCC 1674

RESULT 7
US-09-712-363-143
; Sequence 143, Application US/09712363
; Patent NO. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS

```

```

; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 143
; LENGTH: 3285
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; US-09-712-363-143

Alignment Scores:
Pred. No.: 1.93 Length: 3285
Score: 97.50 Matches: 66
Percent Similarity: 30.91% Conservative: 19
Best Local Similarity: 24.00% Mismatches: 95
Query Match: 8.98% Indels: 96
DB: 9 Gaps: 12

US-09-809-545A-2 (1-203) x US-09-712-363-143 (1-3285)

Qy 1 MethrAsnLysLysAlaValAsnProTyrThrAsnGlyTTrpLysLeuAnProValVal 20
Db 2203 ATGGCGAGCGACGTCGTGGCGGACCCGACCCCAATGCCGCGGATGCAACCGGTTCCG 2262

Qy 21 GlyAlaValTyrSerProAspPheTyrAlaGlyThrValLeuLeuLysGlnAlaAsnGln 40
Db 2263 GCGCAGCGTTCGACCGAGCGACCGCTGGCGGTATCATCCGTCGCTCAAAACC 2322

Qy 41 Glu-----GlySerSerMetTyrSerGlyPro-----SerSerLeuValTyr 54
Db 2323 GAGGCGGTGGGCGAGACCTCAAGTCCGACCCGCGGTGCTCCAAACCGCGGCTGTCMAC 2382

Qy 55 ThrSerAlaMetProGlyPheProTyrProAlaAlaThrAlaAlaAlaTyrArgGly 74
Db 2383 TCCATGCGTGGCCCAACCAACCGCGCATCACTCCGCGGCGACCGCGCGA 2442

Qy 75 AlaHisLeuArgGlyArgGlyArgThrValTyrAsnThrPheArgAlaAlaAlaPro--- 93
Db 2443 -----GGAGAGCGCGCGTGGGATCAACGGGTGCAACGCGCGCTGCTTC 2490

Qy 94 -----ProProIleProAlaTyrGly----- 101
Db 2491 GGATTGACCCGCGACGATACCCCGGTGATGGGACGATACGGGAGAACCACTGGCGCC 2550

Qy 102 -----GlyValValTyrGlnGluPro----- 108
Db 2551 ACGCCACCTCGCGCTGATCAAGTATACCGCCCGCAGCCGCGACCGCGCTGCTGTG 2610

Qy 109 -----ValTyrGlyAsn 112
Db 2611 GTTTCGCGCGCGCGCATCTGATCTTCAAGAGAGACGCGCGATTCATTAAGGCGAG 2670

Qy 113 LysLeu-----LeuGlnGlyGlyTyrAlaAlaTyrArgTyrAlaGlnProThrProAlaThr 131

```

Db	2671	TCCTGAAACGACAGTg-GGG-----CGTCAACGGCCCGACGGCCGATCAACCACT	2723
Qy	132	AlAlAlAlaIaIySerAspSerTyI-----S-----	139
Db	2724	GGGGCGAGGTATTTCGATCGACATCGGACCGCAACCCGGTGGCGAGACTCGGGTTCC	2783
Qy	140	-----GIaIyGValIyTAlaIaAlaAspProTyrHisIa-----	150
Db	2784	GCTGGCCTGGGCGCGCCGCGAGGGCCGACGTGGCGC-----CATGTGCG	2828
Qy	151	-----ThIeulAlaProAlaAspProThIyGly	159
Db	2829	CTATGACCCCGAAGCTAGACCCCTGAGCAATGTCGCTTCAACCCCGCGGGGTTCCGT	2888
Qy	160	ValGIyAlaMetAlaAlaAlaAlaAlaProLeuThIyAspAlaAluSerThIyArgSerHisAlaAsp	179
Db	2889	GCTGGAAATCTCTGACGGCGGTTGATCGGGTCAGACGACACGGGTGGATGGACATCGCAC	2948
Qy	180	AspValGIyLeuValLeuSerSerLeuGAlaAlaSerIleTyrGln	194
Db	2949	CGCAGCCAACTT-----CCCTTCACGACCAACCGCTTTTCGA	2994

```

RESULT 8
US-08-754-311B-1
: Sequence 1, Application US/08754311B
: Patent No. US20020004221A1
: GENERAL INFORMATION:
: APPLICANT: Bonini, Nancy M.
: APPLICANT: Leiserson, William M.
: APPLICANT: Benzert, Seymour
: TITLE OF INVENTION: PROGRAMMED CELL DEATH ANTAGONISTS
: TITLE OF INVENTION: PROTEINS
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESSES:
: ADDRESSES: Flahr, Hobach, Test, Albritton & Herbe
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: United States
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/754, 311B
: FILING DATE: 21-NOV-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/195,152
: FILING DATE: 14-FEB-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Treccarlin, Richard F.
: REGISTRATION NUMBER: 31,801
: REFERENCE/DOCKET NUMBER: A-59551/RPT/RMS
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TEXT: 910 277299
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3231 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: US-08-754-311B-1

Alignment Scores:
Pred. No.: 2.66 Length: 3261
Score: 96.00 Matches: 45
Percent Similarity: 37.57% Conservative: 23
Best Local Similarity: 24.86% Mismatches: 63

```

Query Match:	8.84%	Indels:	51
DB:	8	Gaps:	6

US-09-809-545A-2 (1-203) X US-08-754-311B-1 (1-3231)

Qy 42 GlySerSerMetTyr-----SerGlyProSerSerSerLeu 52

Db 1062 GGATCCAATTGTACGGCTGCAGCTCGGCCAGCAATCCGCTGGACGGAGGACAGTGGCG 1121

53 VAlTyrThrSerAlaMetProGlyPheProItyrProAlaAlaIhfrAlaAlaAlaItyf /2

DB 1122 G1CAAC1C11CGCAG1G-----GCAAGCGGCACACACACACCG1C1AC 110

[illegible]

11

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

113 IwstIeulEuglNgIvGIVtVrA]aA]aTvArCtVrA]agI]nProThrProA]aThrA]a 132

Db 1272 GCCAGATGGAACCCGAGCGGCGAGCTGCGGCGGCTGCTACTTGACGCCCAGCTATGCC 1331

QY 133 A1a1a----- 134

Db 1332 GCCAGCGGCAACAACTCGCAGCTGTACAGCAGTCCGTACGCCGGCTACAACTTC 1391

QY 135 -----TyrSerAspSertyrGlyargvalTyrAlaAlaAsp 146

Db 1392 GGGCAGCAGGACTACGGCGGCTACTACAACGAGCAGTACGGCACTATTACAGTCCGGCC 1451

147 ProtyrH1SH1STHrLeuAlaProAlaPro-----InrIyG1yValGly--- 161

DB 1432 AACIACI CACCGIAI GCGI CAGC I CCGCCAGC I CGAG I CGAG I CAI GAGC I GGC I AC 101

100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611

1. *Introduction*

$\frac{1}{\sqrt{2}}$

RESULT 9

US-09-974-300-1655
: Sequence 1655, Application US/09974300

; FACILEC NO.: 0520020148/2171
; GENERAL INFORMATION:

APPLICANT: Clausen, Ib Groth

	TITLE OF INVENTION:	Expression
INT D REFERENCE NO.	NO OF PAGES	FIGS TO

; CURRENT APPLICATION NUMBER: US/09/974,300
 ; CURRENT FILING DATE: 2001-10-05

PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 09/680,598
;

; PRIOR APPLICATION NUMBER: 80/219,328
 ; PRIOR FILING DATE: 2001-03-27

```
; SOFTWARE: FastSeq for Windows Version 4.0
```

LENGTH: 657

ORGANISM: *Bacillus licheniformis*

```

; NAME/KEY: misc_feature
; LOCATION: (1) (657)

```

OTHER INFORMATION: n = A, I, C or G
US-09-974-300-1655

Alignment Scores:
Pred. No.: 0.569 Length: 657
Score: 94.00 Matches: 38
Percent Similarity: 40.15% Conservative: 17
Best Local Similarity: 27.74% Mismatches: 58
Query Match: 8.65% Indels: 24
DB: 10 Gaps: 8

US-09-809-545a-2 (1-203) x US-09-974-100-1655 (1-657)

QY 2 ThrAsnLysLysAlaValAsnProTyrThrAsnGlyTTPlyLysLeuAsnProValValGly 21
Db 100 ACAAAAAAAAAACCAATTAAACCA---ACGAGGGCTTAAACGAAACGGTAAACTCC 156
QY 22 -----AlaValTyrSerProAspPhe---TyrAlaGlyThrValLeuLeu 35
Db 157 TTAACCGCGCCCGCGGCTTTTCCAAGATTTCTTATACAAAACCTGGGCTCTT 216
QY 36 CysGlnAlaAsnGlnGlySerSerMetTyrSerGlyProSerSerLeuValTyrThr 55
Db 217 ACCACCAAAAAAACAAGGTCGGGTTTC-----CCCTTAAAGAAAAATTAAAC 267
QY 56 SerAlaMetProGlyPheProTyrProAlaAlaThrAlaAlaAlaAlaTyrArgGlyAla 75
Db 268 CCCTTTAAACCAATTCACTTTTAACTTAAACGCAACCCCAATCCCAACCCCTC 327
QY 76 HisLeuArgGlyArgGlyArgThrValTyrAsnThrPheArgAlaAlaAlaProProPro 95
Db 328 CATATTTCATCCTTGGT-----TTCAAAAATTTTTCATTCATTTAAACCCCTCG 378
QY 96 ProIleProAla-----TyrGlyGlyValValTyrGlnGln 107
Db 379 AAAAACCCCTCCCGGTACAACTATTATTTTCAATATCAACAGGGGCTTGGCTCGGAA 438
QY 108 ProValTyrGlyAsnLysLeuLeuGlnGly-----TyrAlaAlaTyr 122
Db 439 CCAATTAACCCCAAC---CTGGCCAAAGCGCGGCTTACTCCTTAC 486

RESULT 10

US-09-822-830A-410/C
Sequence 410, Application US/09822830A
Patent No. US20020142952A1
GENERAL INFORMATION:
APPLICANT: Genetics Institute, Inc.
APPLICANT: Wong, Gordon G.
APPLICANT: Clark, Hilary
APPLICANT: Fechtel, Kim
APPLICANT: Agostino, Michael J.
APPLICANT: Howes, Steven H.
APPLICANT: Resnick, Richard J.
APPLICANT: Gulukota, Kamalakkar
APPLICANT: Graham, James R.
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
FILE REFERENCE: GIN 6402
CURRENT APPLICATION NUMBER: US/09/822, 830A
CURRENT FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/195,604
PRIOR FILING DATE: 2000-04-06
NUMBER OF SEQ ID NOS: 631
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 410
LENGTH: 2273
TYPE: DNA
ORGANISM: Homo sapiens
US-09-822-830A-410

Alignment Scores:
Pred. No.: 3.8 Length: 2273
Score: 92.50 Matches: 48
Percent Similarity: 33.72% Conservative: 10
Best Local Similarity: 27.91% Mismatches: 73
Query Match: 8.52% Indels: 41
DB: 10 Gaps: 9

US-09-809-545a-2 (1-203) x US-09-822-830A-410 (1-2273)

QY 29 TyrAlaGlyThrValLeuLeuLeuCysGlnAla---AsnGlnGlySerSerMetTyrSer 47
Db 2182 TATGCACAGACACAC-----CAGGCATATGGCAACCAAGCTATGAACTATGCA 2132
QY 48 GlyProSerSerLeuValTyrThrSerAlaMetProGlyPheProTyrProAlaAlaThr 67
Db 2131 CAGCCCATGATGTGACGTATATCCAGGCTCAGACCACTGCACCTATGGCAGACGCC 2072
QY 68 AlaAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArgGlyAlaGlnThrValTyrAsnThr 87
Db 2071 TATCAACTTCTTAT-----GGACAGCCTCCACATGCTTAATACTACT 2030
QY 88 PheArgAlaAlaAlaProProProProIleProAlaTyrGlyGlyValValValTyrGlnGln 107
Db 2029 -----CNACTGCCCCCCAG-----GCATACAGCCAG 2003
QY 108 ProValTyrGlyAsnLysLeuLeuGlnGlyGlyTyrAlaAlaTyrArgTyrAlaGlnPro 127
Db 2002 CCGTCCAGGGGTAT-----GGCACTGGTGCTTATGATACCAACCACTGCT 1958
QY 128 ThrProAlaThrAlaAlaAlaTyrSerAspSerTyrGlyArgValTyrAlaAlaAspPro 147
Db 1957 ACAGTCACCAACCCAGGCC-----TCTATGCACTCAGTCT 1919
QY 148 TyrHisHisThrLeuAlaProAlaProThrTyrGlyValGlyAlaMetAsnAlaPheAla 167
Db 1918 GCATATGGCACTCAGCCTGCTTATTCAGCTATGGGAG-----CAG 1877
QY 168 ProLeuThrAspAlaLysThrArgSerHisAlaAspValGlyLeuValLeuSerSer 187
Db 1876 CCAGCAGCACTGACACCTACAGACAGCCGAGATGGAACCAAGCCACTGAGACTAGTCAA 1817
QY 188 LeuGlnAlaSerIleTyrGlnGlyGlyTyrAsnArg 199
Db 1816 CCTCAATCTAC-----ACAGGGGGGTACCAACAG 1787

RESULT 11

US-09-880-107-3769
Sequence 3769, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3769
LENGTH: 2390
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 X66899
US-09-880-107-3769

Alignment Scores:
Pred. No.: 4.05 Length: 2390
Score: 92.50 Matches: 48
Percent Similarity: 33.72% Conservative: 10
Best Local Similarity: 27.91% Mismatches: 73
Query Match: 8.52% Indels: 41
DB: 10 Gaps: 9

PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 1143
LENGTH: 1782
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-1143

Alignment Scores:
Pred. No.: 4.94 Length: 1782
Score: 90.00 Matches: 57
Percent Similarity: 30.50% Conservative: 22
Best Local Similarity: 22.01% Mismatches: 66
Query Match: 8.29% Indels: 114
DB: 9 Gaps: 12

US-09-809-545A-2 (1-203) x US-09-938-842A-1143 (1-1782)

QY 5 LyeAlaValAsnProTyrThrAsnGlyTTrpLysLeuAsnProValValGlyAlaValTyr 24
DB 163 CGATCTTTCTCACTTATGCTTACCCAGTATGCTTGATCTTCTCTGTTTCTAT 216
QY 25 SerProAspPheTyrAla-----GlyThrValLeuLeuCysGln 37
DB 217 GTTCCCTAATGTGATATCAGACGCCCTACTATTATGTTATGGAGTGTAGTACACAGGTTAT 276
QY 38 AlaAsnGlnGlySerSerMetTyrSerGlyPro-----SerSerLeuVal 53
DB 277 ACCAATTCCGAGTGTGATATGATCATCATCAGAGCTTACGAGAGAAATGATCTCTGTT 336
QY 54 TyrThrSer-----AlaMetProGlyPheProTyrProAlaThrAlaAlaAla 71
DB 337 TATCCGACAGGATACGGGATATGCTGCTTCTTCTTACTCACACAGCAAGCCCTGCTCA 396
QY 72 TyrArgGly-----AlaHisLeuArgGlyArgGlyThrValTyr----- 85
DB 397 CAACCTGGTGGTATGATGACAGTGTATGGTGTACAGACATACAGTATCTTCTCTCTC 456
QY 86 -----Asn 86
DB 457 ACAGCCAGTAGTGGCCCTTTCCTGCTGCTGCTTCTGCTTACTGACAGCAAGCTCTCT 516
QY 87 ThrPheArgAlaAla-----Ala 92
DB 517 ACAACAAAGCAGCCAACTCTGCATCTGCTGTATCCCTAAAGATGATGATCAGCT 576
QY 93 ProProProProLeuProAlaTyrGlyValValTyrGlnGlnProValTyrGlyAsn 112
DB 577 CCGTTAAACCATG-----AATCAGAGTGAATGTATGGAAC 615
QY 113 LysLeuLeuGlnGlyTyrAlaAlaTyrArgTyrAlaGlnPro----- 127
DB 616 AGTGCCTTGAAGTGTCTTCTGCTGCTGCTTATCAGAGCCAGTATAGTATGAT 672
QY 128 -----ThrProAlaThr----- 131
DB 673 GGGTTTACACTCTGCTGCTGATGAGCAGATGCTTACTTCAAGTGCAGAGGCT 732
QY 132 -----AlaAlaAlaTyrSerAspSerTyrGlyArgVal----- 142
DB 733 GTTTCGGCAGTGGAGTGTATCTATATTCTAAGCTTAACAACATGTATCCTGCACT 792
QY 142 ----- 142
DB 793 AGAATCAAACTCTAGCTCAATTCGATTAACAGAGTATGACAGCCCTGCATCCATG 852
QY 143 -----TyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGly 159
DB 853 ACAGTTATGCACTCAGGGTGTACTACGACAGAGTGTCTCCAAACAGTATATGGT 909

Search completed: March 16, 2003, 04:04:03
Job time : 122 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2003, 21:20:45 ; Search time 40 Seconds
(without alignments)
676.247 Million cell updates/sec

Title: US-09-809-545A-2

Perfect score: 1086

Sequence: 1 MTNKAIVNPTNGMKLNPPV.....VLSLSIASIYOGYRFPAPY 203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1086	100.0	203	AAU70146	Rat secreted facto
2	1009.5	93.0	397	AA892573	Human protein sequ
3	1009.5	93.0	440	AAW79112	Human protein SEQ
4	551	50.7	330	AAV81462	Human ataxin-2 bin
5	534	49.2	135	ABG02058	Novel human diagno
6	534	49.2	237	AAW40093	Human polypeptide
7	534	49.2	268	AAW40092	Human polypeptide
8	405.5	37.3	705	ABG29447	Novel human diagno
9	361	33.2	291	AAW41878	Human polypeptide
10	361	33.2	291	AAW41879	Human polypeptide

11	342	31.5	102	22	ABG29445	Novel human diagno
12	278	26.5	366	21	AAV56850	Human RNA binding
13	278	25.6	388	22	ABG26323	Novel human diagno
14	258	23.8	286	22	AAW80096	Human protein SEQ
15	226	20.8	101	22	ABG29446	Novel human diagno
16	178	16.4	151	22	ABG26319	Novel human diagno
17	129	11.9	367	22	ABG29047	Human polypeptide
18	127.5	11.7	123	22	AAW06790	Human polypeptide
19	116	10.7	3063	23	ABW9762	Human Tumour Endot
20	116	10.7	3118	22	AAU27790	Human full-length
21	110.5	10.2	219	22	ABW60573	Microsomopora eve
22	110	10.1	474	22	AAU04831	Human polypeptide
23	108.5	10.0	244	22	AAW38864	Human polypeptide
24	108.5	10.0	297	22	AAW40650	Human protein SEQ
25	103.5	9.5	939	22	AAW78860	Human protein SEQ
26	103.5	9.5	943	22	AAW79844	Human protein SEQ
27	101.5	9.3	950	22	ABW71513	Drosophila melanog
28	100	9.2	830	22	ABW59880	Drosophila melanog
29	98.5	9.1	655	23	ABW57126	Mouse ischaemic co
30	98	9.0	376	22	ABW59035	Drosophila melanog
31	97.5	9.0	828	22	ABW58368	Drosophila melanog
32	97	8.9	179	22	ABW67727	Drosophila melanog
33	96.5	8.9	748	23	ABW92874	Herbicidally activ
34	96	8.8	388	23	ABG53243	Human prostate spe
35	96	8.8	760	18	AAW29490	Programmed cell de
36	96	8.8	760	23	AAU75570	Programmed cell de
37	96	8.8	766	22	ABW64516	Drosophila melanog
38	94	8.7	349	21	AAW5691	Murine Sam68-like
39	94	8.7	349	22	AAW49332	Murine SLM-1-prote
40	94	8.7	388	23	ABG61877	Prostate cancer-as
41	94	8.7	576	22	ABW61241	Drosophila melanog
42	93	8.6	444	19	AAW41940	Corn p-hydroxyphen
43	93	8.6	2897	22	ABW58514	Drosophila melanog
44	92.5	8.5	187	22	ABW67725	Drosophila melanog
45	92.5	8.5	362	18	AAW34972	Chimeric Ewing's s

ALIGNMENTS

RESULT 1	
AAU70146	
AAU70146 standard; Protein; 203 AA.	
XX	
AC AAU70146;	
XX	
DT 12-MAR-2002 (first entry)	
XX	
DE Rat secreted factor protein encoded by DNA clone P0184_D11.	
XX	
KW Rat; secreted factor polypeptide; cardiac disease; renal disease; kidney;	
KW inflammatory disease; congestive heart failure; myocarditis; asthma;	
KW dilated congestive cardiomyopathy; angina pectoris; cardiac arrhythmia;	
KW myocardial infarction; pulmonary hypertension; arteriosclerosis; stroke;	
KW atherosclerosis; cardiac tumour; glomerulonephritis; nephrotic syndrome;	
KW renal infarction; hereditary nephritis; polycystic kidney disease;	
KW chronic renal failure; renal vein thrombosis; medullary sponge kidney;	
KW rheumatoid arthritis; osteoarthritis; psoriasis; restenosis;	
KW graft versus host reaction; Crohn's disease; ulcerative colitis;	
KW Alzheimer's disease; gene therapy.	
XX	
OS Rattus norvegicus.	
XX	
PN WO200174901-A2.	
XX	
PD 11-OCT-2001.	
XX	
PF 23-MAR-2001; 2001WO-US09555.	
XX	
PR 31-MAR-2000; 2000US-193548P.	
XX	
PA 14-MAR-2001; 2001US-0809545.	
XX	
PA (SCIO-) SCIOS INC.	

PI Stanton LW, White RT;
 XX
 XX WPI: 2002-010779/01.
 DR N-PSDB; AAS94693.
 XX
 PT Novel secreted factor polypeptide useful for treating cardiac diseases
 PT such as arteriosclerosis, myocardial infarction, inflammatory diseases
 PT such as asthma, stroke, and rheumatoid arthritis and renal diseases -
 XX
 PS Claim 18, Fig 1, 189pp; English.
 XX
 CC The invention relates to rat secreted factor polypeptides and the
 CC polynucleotides encoding them. The sequences are useful for treating
 CC cardiac, renal or inflammatory diseases. These include cardiac diseases
 CC such as congestive heart failure, myocarditis, dilated congestive
 CC cardiomyopathy, angina pectoris, myocardial infarction, cardiac
 CC arrhythmia, pulmonary hypertension, arteriosclerosis, atherosclerosis and
 CC cardiac tumours, renal diseases such as glomerulonephritis, nephrotic
 CC syndrome, renal infarction, hereditary nephritis, polycystic kidney
 CC disease, chronic renal failure, renal vein thrombosis and medullary
 CC sponge kidney and inflammatory diseases such as asthma, rheumatoid
 CC arthritis, osteoarthritis, stroke, psoriasis, xeroderma, grapt versus
 CC host reaction, Crohn's disease, ulcerative colitis and Alzheimer's
 CC disease. Sequences AAU70146-AAU70178 represent the secreted factor
 CC polypeptides of the invention.

PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241859.
 XX
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Oka T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wamamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 XX
 PR Primer sets for synthesizing polynucleotides, particularly the 5602
 PR full-length cDNAs defined in the specification, and for the detection
 PR and/or diagnosis of the abnormality of the proteins encoded by the
 PR full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 10787; 2537pp + CD ROM; English.

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US04098.
 XX
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 XX
 PA (HXSE-) HXSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX
 DR WPI; 2001-476283/51.
 DR N-PSDB; AAK52245.
 XX
 PT Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX
 PS Claim 20; Page 4126-4127; 6221pp; English.
 XX
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAW7832-AAW80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAW80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 XX
 SQ Sequence 440 AA;
 Query Match 93.0%; Score 1009.5; DB 22; Length 440;
 Best Local Similarity 93.6%; Pred. No. 4.1e-88;
 Matches 190; Conservative 4; Mismatches 8; Indels 1; Gaps 1;
 QY 1 MTNKAANPNTGKLNPNVGVSPDPAAGTVLLCOANQDESSMSYSGSSLVTSAMPG 60
 DB 229 MTNKTVPNTGKLNPNVGVSPDPAAGTVLLCOANQDESSMSYSAFSSLVTSAMPG 298
 QY 61 PPYPAAATAAARGAHLRGGRGVNTTPRAAPPPPIPAVGWVVOEPYGNKLLQGGYA 120
 DB 299 PPYPAAATAAARGAHLRGGRGVNTTPRAAPPPPIPAVGWVVOEGYGD-ITYGYA 357
 QY 121 AYRYAQPPTAAAYSDSYGRVYAADPYNHTLAPATYGVGMANAFAPLTDAKTRSHADD 180
 DB 358 AYRYAQPPTAAAYSDSYGRVYAADPYNHTLAPATYGVGMANAFAPLTDAKTRSHADD 417
 QY 181 VGLVSSLSQASLYOGGYNRFAPY 203
 DB 418 VGLVSSLSQASLYOGGYNRFAPY 440

RESULT 4
 AAY81462
 ID AAY81462 standard; Protein; 330 AA.
 XX
 AC AAY81462;
 XX
 DT 03-JUL-2000 (first entry)
 XX
 DE Human ataxin-2 binding protein (A2BP).
 XX
 KW Ataxin-2 binding protein; A2BP; human; RNA-binding; cell polarisation;
 KW neuronal plasticity; cellular degeneration signal transduction pathway;
 KW selective RNA transport; spinocerebellar ataxia type-2;
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Region 1023..1312
 FT /note="This region contains the regions necessary
 FT for binding to ataxin-2"
 XX
 PN WO200012710-A1.
 XX
 PD 09-MAR-2000.
 XX
 PF 01-SEP-1999; 99WO-US20156.
 XX
 PR 01-SEP-1998; 98US-0145381.
 XX
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.
 XX
 PI Pulst SM, Shibata H;
 XX
 DR WPI; 2000-237873/20.
 DR N-PSDB; AAA07075.
 XX
 PT Nucleic acids encoding an ataxin-2 binding protein useful for
 PT inhibiting the expression of active proteins from the SCA2 gene for the
 PT treatment of spinocerebellar ataxia type-2 -
 XX
 PS Claim 19; Page 77-78; 82pp; English.
 XX
 CC This sequence represents cDNA encoding human ataxin-2 binding protein
 CC (A2BP). Nucleotide sequences encoding human A2BP were originally
 CC isolated in an adult brain cDNA library using the yeast two hybrid
 CC method. The ligand of A2BP, ataxin-2, is a 40 kD protein of unknown
 CC function that is encoded by the SCA2 gene located on chromosome 12. SCA2
 CC has been linked to the autosomal dominant neurodegenerative disorder
 CC spinocerebellar ataxia type-2. Individuals afflicted with the disease
 CC exhibit CAG triplet expansion in the SCA2 gene, resulting in ataxin-2
 CC containing a polyglutamine stretch of about 35-39 residues, whereas that
 CC of normal individuals contains approximately 22 contiguous glutamine
 CC residues. A2BP and ataxin-2 are components of a cellular degeneration
 CC signal transduction pathway. The pathogenic expanded form of ataxin-2 has
 CC a higher affinity for A2BP relative to normal ataxin-2; the presence of
 CC the expanded form is likely to promote degeneration. A2BP and ataxin have
 CC also been found to have a role in gene regulation. The binding of A2BP to
 CC ataxin-2 plays an important role in controlling gene expression via the
 CC targeting of transport of specific RNAs, selective RNA transport being
 CC mediated via the RNA binding domains of A2BP. A2BP is expressed very
 CC early in embryonic development. Both ataxin-2 and A2BP are able to bind
 CC RNA, and are essential components of the RNA localisation network that
 CC establishes cellular polarity in embryogenesis. In highly differentiated,
 CC polarised cells such as neurons, A2BP and ataxin-2 have a similar
 CC function and are required for neuronal plasticity. A2BP nucleic acids may
 CC be used for the recombinant production of A2BP proteins or fragments
 CC thereof according to standard methodologies. For example, an A2BP protein
 CC with an ataxin-2 or RNA binding capability but no signal transduction
 CC function can be used as a dominant negative inhibitor of the cellular
 CC degeneration signal transduction pathway. A2BP proteins with a signal
 CC transduction function can be used to treat hyperproliferative disorders

CC (e.g., cancer) via stimulation of the cellular degeneration pathway.
XX
SQ Sequence 330 AA;

Query Match 50.7%; Score 551; DB 21; Length 330;
Best Local Similarity 97.1%; Pred. No. 2e-44; 2; Indels 0; Gaps 0;
Matches 102; Conservative 1; Mismatches 2;

Oy 1 MTNKA VNPVTGKMLNPNVGA VSPDPFA GTVLLCQANQEGSSMYSGPSSLVYTSAMPG 60
Db 196 MTNKKTVNVTYTGKMLNPNVGA VSPDPFA GTVLLCQANQEGSSMYSGPSSLVYTSAMPG 255

Oy 61 PPYPATAAAAYRG AHLRGRGRTVNTFRAAAPPPPIPAVGGVY 105
Db 256 PPYPATAAAAYRG AHLRGRGRTVNTFRAAAPPPPIPAVGGVY 300

RESULT 5
ABG02058
ID ABG02058 standard; Protein; 135 AA.

AC ABG02058;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #2049.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YF;

DR WPI: 2001-639362/73.

DR N-PSDB; AAS66245.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 20; SEQ ID NO 32417; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 135 AA;

Query Match 49.2%; Score 534; DB 22; Length 135;
Best Local Similarity 64.0%; Pred. No. 2.7e-43;
Matches 105; Conservative 6; Mismatches 23; Indels 30; Gaps 4;

Oy 1 MTNKA VNPVTGKMLNPNVGA VSPDPFA GTVLLCQANQEGSSMYSGPSSLVYTSAMPG 60
Db 1 MTNKKTVNVTYTGKMLNPNVGA VSPDPFA GTVLLCQANQEGSSMYSGPSSLVYTSAMPG 255

Oy 61 PPYPATAAAAYRG AHLRGRGRTVNTFRAAAPPPPIPAVGGVYQEPYGNKLLGGYA 120
Db 34 PPYPPTGTAVAYRG AHLRGRGRTVNTFRAAAPPPPIPAVGAVYQDGYAE-IYGVYA 92

Oy 121 AVRYAQPPTATTAAYSDSYGRVY-AADPYHHTLAPPTYGVA 163
Db 93 AVRYAQPFA-AAAAYSDSYGRVYAADPYHHTIGPATYSIGTM 135

RESULT 6
AAM40093
ID AAM40093 standard; Protein; 237 AA.

AC AAM40093;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 3238.

KW Human; nocotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.

OS Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

PA (HYSE-) HYSEQ INC.

PI Tang YF, Liu C, Auand V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AU, Yang Y, Zhang J;
PI Zhao Q, Zhou P, Goodrich R, Drmanac RT;

DR WPI: 2001-442253/47.

DR N-PSDB; AAI59249.

XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 5; SEQ ID NO 3238; 10078pp; English.

CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nocotropic,

CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Actin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
CC
SQ Sequence 237 AA;
Query Match 49.2%; Score 534; DB 22; Length 237;
Best Local Similarity 64.0%; Pred. No. 5.5e-43;
Matches 105; Conservative 6; Mismatches 23; Indels 30; Gaps 4;
QY 1 MTNKKAVNPYTNMGKLNPNVGAIVSPDFYAGTVLLCOANQCGSSMYSGPSSLVYTSAMP 60
DB 103 MTNKKGNPPTNGKMLPNVGAIVGPEFYAVT-----G 135
QY 61 FPPYPAATAAAYRGAGHLRGRGRTVYNTFRRAAPPPIPAYGVVYOEPVGNKLLQGGYA 120
DB 136 FPPYTTGTAAVAYRGAGHLRGRRAVYNTFRRAAPPPIPTYGAVVYQDGFYGAE-IYGGYA 194
QY 121 AYRYAOPTPATAAAYSDSYGRVY-AADPYHHTLAPATYGVGM 163
DB 195 AYRYAOPA-AAAAAYSDSYGRVYAAADPYHHTIGPATYISIGTM 237
RESULT 7
AAM40092
ID AAM40092 standard; Protein: 268 AA.
XX AAM40092;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 3237.
XX
KM Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KM peripheral nervous system; neuropathy; central nervous system; CNS;
KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KM chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KM leukaemia.
XX
OS Homo sapiens.
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000MO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0582317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J,
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.
DR N-PSDB; AAI59248.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 5; SEQ ID NO 3237; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic;
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Actin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
CC
SQ Sequence 268 AA;
Query Match 49.2%; Score 534; DB 22; Length 268;
Best Local Similarity 64.0%; Pred. No. 6.5e-43;
Matches 105; Conservative 6; Mismatches 23; Indels 30; Gaps 4;
QY 1 MTNKKAVNPYTNMGKLNPNVGAIVSPDFYAGTVLLCOANQCGSSMYSGPSSLVYTSAMP 60
DB 134 MTNKKGNPPTNGKMLPNVGAIVGPEFYAVT-----G 166
QY 61 FPPYPAATAAAYRGAGHLRGRGRTVYNTFRRAAPPPIPAYGVVYOEPVGNKLLQGGYA 120
DB 167 FPPYTTGTAAVAYRGAGHLRGRRAVYNTFRRAAPPPIPTYGAVVYQDGFYGAE-IYGGYA 225
QY 121 AYRYAOPTPATAAAYSDSYGRVY-AADPYHHTLAPATYGVGM 163
DB 226 AYRYAOPA-AAAAAYSDSYGRVYAAADPYHHTIGPATYISIGTM 268
RESULT 8
ABG29447
ID ABG29447 standard; Protein: 705 AA.
XX ABG29447;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #29438.
XX
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001MO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.
DR N-PSDB: AAS93634.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 20; SEQ ID NO 59606; 103bp; English.

XX
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostic production of (II). The
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 705 AA;

Query Match 37.3%; Score 405.5; DB 22; Length 705;

Best Local Similarity 34.4%; Pred. No. 4,4e-30; Indels 181; Gaps 7;
Matches 105; Conservative 9; Mismatches 10;

QY 53 VYTSMPGPPYATTAAYRGALRGRTVYNTFRRAAPPPIPAVGCV----- 103
DB 190 VEINSVPGFPYATTAAYRGALRGRTVYNTFRRAAPPPIPAVGGLNVLDPY 249
QY 104 -----VY--OE-----PVYG 111
DB 250 FALSDDDFNILLVCIRVYTGQOETSLCHCPAQCKRDLQKMFVDAQCPQLGSPV-- 307
QY 112 NKLLDGGVAAVYAOPTATTAAYSP----- 137
DB 308 PRLGGGVAAYRAOPTATTAAYSDRLSLMGDSSEIVKGRNQFPVAADDEISCNTSA 367
QY 138 ----- 137
DB 368 GCIPTVQLEASSGVGCVTSAPHSIFPTSQPTHAASYIKTLDSGDAMADKIDTAPASMEF 427
QY 138 -----SYGRVYAADPYHHTLAAPPTYGVGAMNAFPLTDAKTRSHADVGLVSLQAS 191
DB 428 SFOEGSSVGRVYAADPYHALAPPTYGVGAMD-----RQY-----SSLCVS 469
QY 192 IYQGG 196
DB 470 VCTGG 474

RESULT 9

AAAM41878
ID AAAM41878 standard; Protein; 291 AA.

AC AAAM41878;

DT 22-OCT--2001 (first entry)

DE Human polypeptide SEQ ID NO 6809.

KW Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
KW peripheral nervous system neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.

OS Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0633450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang Z, Wehrman T, Xu C, Xue AD, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.

XX N-PSDB: AAI61034.

XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -

XX Example 2; SEQ ID NO 6809; 10078bp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
XX the encoded polypeptides (AAAM3642-AAAM4213) with nootropic,
XX immunosuppressant and cyostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.

SO Sequence 291 AA;

Query Match 33.2%; Score 361; DB 22; Length 291;

Best Local Similarity 55.5%; Pred. No. 2.5e-26; Indels 34; Gaps 2;
Matches 71; Conservative 5; Mismatches 18;

QY 1 MTNKAANDPYTNGMKTLPVGAIVSPDPYAGTVLLCQANQEGSSMTSGSLVYTSAMPG 60

DB 147 MTNKKTGNDPYTNGMKTLPVGAIVGPEFYAVT-----G 179

QY 61 FPYPAATTAAYRGALRGRTVYNTFRRAAPPPIPAVGWVYQEPYGNKLLGGYA 120

DB 180 FPYPTTGAVAVRGALRGRTVYNTFRRAAPPPIPIYGAIVVIOGRTGAEILEA--- 236

QY 121 AYRYAOPT 128

DB 237 ----TQPT 240

Query Match	Best Local Similarity	Score	DB	Length
Matches 71; Conservative 5; Mismatches 18; Indels 34; Gaps 2;				

1 MTNKAIVNYYTGMKLNIPVGAIVSGDFPAGTVLLCOANQDESSMYSGPSISLVYISAMPG 60

Db 147 MTNKTGVPYTNKMLPVGAVYGPEFTAVT-----G 179

Qy 61 FFPAPATAAAARAHNRGRGTVTNTFFRAAAPPPPIPAVGVWVOEPYNGKLLOGGYA 120
180 FFYPITGTAIVAGRAHNRGRGRAVYNTFFRAAPPPIPPIPGAVVYDDRGYAEILEA--- 236

Db 121 AVRYAQPT 128
237 ----TOPT 240

RESULT 11

ABG29445 ID ABG29445 standard; Protein; 103 AA.

AC ABG29445;

XX 18-FEB-2002 (first entry)

DT DE Novel human diagnostic protein #29436.

XX XX Human: Chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
OS MOZ00175067-A2.
PN 11-OCT-2001.
PD 30-MAR-2001; 2001WO-US08631.
PF 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
PR (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
PI WPI: 2001-639362/73.
XX N-PSDB; AAS93632.
DR

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
PT

XX Claim 20; SEQ ID No 59804; 103bp; English.

PS The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (III) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.
XX Sequence 103 AA:

Query Match 31.5%; Score 342; DB 22; Length 103;
 Best Local Similarity 97.0%; Pred. No. 4.3e-25;
 Matches 64; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 138 SYGRVYADPHTHTAPPTVGVGMNAPFLTDKTRSHADVDGLVLSLQASITVGGY 197
 DB 4 SYGRVYADPHTHTAPPTVGVGMNAPFLTDKTRSHADVDGLVLSLQASITVGGY 63

OY 198 NRPAPY 203
 DB 64 NRPAPY 69

RESULT 12
 ID AAY56850 standard; Protein; 366 AA.
 XX
 AC AAY56850;
 XX
 DT 11-APR-2000 (first entry)
 XX
 DE Human RNA binding protein (RNABP)-2 (clone 1250374).
 XX
 KM RNA binding protein; RNABP; cancer; immune disorder; AIDS; human;
 KM developmental disorder; acquired immunodeficiency syndrome; RNABP-2;
 KM inflammation; allergy; diabetes mellitus; seizure disorder.
 XX
 OS Homo sapiens.
 XX
 PN US620164-A.
 XX
 PD 01-FEB-2000.
 XX
 PF 21-OCT-1998; 98US-0176657.
 XX
 PR 21-OCT-1998; 98US-0176657.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Bandman O, Corley NC, Lu DM, Baughn MR, Tang YT, Guegler KJ;
 DR WP1; 2000-146885/13.
 DR N-PSDB; AA246827.
 XX
 PT Isolated and purified polynucleotide for modulating the expression of
 PT human RNA binding proteins which play a role in cancer; immune
 PT disorders and developmental disorders -
 XX
 PS Examples; Column 45-48; 39pp; English.
 XX
 CC The present invention provides human RNA binding proteins (RNABP) and
 CC polynucleotides encoding the proteins. The polypeptides are useful for
 CC modulating the expression of human RNA binding proteins (RNABP) which
 CC play a role in cancer, immune disorders and developmental disorders.
 CC Disorders associated with a decrease of RNABP include: cancers such as
 CC adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma,
 CC teratocarcinoma, and, in particular, cancers of the adrenal gland,
 CC bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia,
 CC gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary,
 CC pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen,
 CC testis, thymus, thyroid, and uterus, immune disorders such as acquired
 CC immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory
 CC distress syndrome, allergies, ankylosing spondylitis, amyloidosis,
 CC anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune
 CC thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's
 CC disease, atopic dermatitis, dermatomyositis, diabetes mellitus,
 CC emphysema, episodic lymphopenia with lymphocytocytosis, erythroblastosis
 CC fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis,
 CC Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis,
 CC hyperosteoarthritis, irritable bowel syndrome, multiple sclerosis,
 CC myasthenia gravis, myocardial or pericardial inflammation,
 CC osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis,

CC Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjogren's
 CC syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic
 CC sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner
 CC syndrome, complications of cancer, hemodialysis, and extracorporeal
 CC circulation, viral, bacterial, fungal, parasitic, protozoal, and
 CC helminthic infections, trauma, X-linked agammaglobulinemia of Bruton,
 CC common variable immunodeficiency (CVID), Diego's syndrome (thymic
 CC hypoplasia), thymic dysplasia, isolated immunoglobulin (Ig)-A deficiency,
 CC severe combined immunodeficiency disease (SCID), immunodeficiency with
 CC thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi
 CC syndrome, chronic granulomatous diseases, hereditary angioneurotic edema,
 CC and immunodeficiency associated with Cushing's disease; and developmental
 CC disorders such as renal tubular acidosis, anemia, Cushing's syndrome,
 CC achondroplastic dwarfism, Duchenne and Becker muscular dystrophy,
 CC epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia,
 CC genitourinary abnormalities, and mental retardation), Smith-Magenis
 CC syndrome, myelodysplastic syndrome, hereditary mucopolysaccharidosis,
 CC hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-
 CC Tooth disease and neurofibromatosis, hypochromidism, hydrocephalus,
 CC seizure disorders such as Sydenham's chorea and cerebral palsy, spina
 CC bifida, anencephaly, craniofacial dysmorphism, congenital glaucoma, cataract,
 CC sensorineural hearing loss, and any disorder associated with cell
 CC growth and differentiation, embryogenesis, and morphogenesis involving
 CC any tissue, organ, or system of a subject, e.g., the brain, adrenal
 CC gland, kidney, skeletal or reproductive system. The present sequence
 CC represents the RNABP-2 polypeptide.
 XX
 SQ Sequence 366 AA;

Query Match 26.5%; Score 288; DB 21; Length 366;
 Best Local Similarity 48.9%; Pred. No. 3.2e-19;
 Matches 68; Conservative 11; Mismatches 54; Indels 6; Gaps 4;

OY 1 MTNKAANPYTNGKLANVGVAVSPDFYAGTLLCOAN--QEGSSMYSGPSSL--VYTS 56
 DB 190 MTNKAANPYTNGKLANVGVAVSPDFYAGTLLCOAN--QEGSSMYSGPSSL--VYTS 249

OY 57 AMPGFPPY-AATTAAYAGAHILRGRTVYTFRAAAPPPIPAVGIVYQEPVGNKLL 115
 DB 250 IIPGFPPYTAATTAAYAGAHILRGRTVYTFRAAAPPPIPAVGIVYQEPVGNKLL 308

OY 116 QGGYAAVRYAOPTPATMAA 134
 DB 309 QPPELLOPLOTVTYVMA 327

RESULT 13
 ID ABG26323 standard; Protein; 388 AA.
 XX
 AC ABG26323;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #26314.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.
DR N-PSDB; AAS90510.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics; forensics; gene mapping; identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID No 56682; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 388 AA;
Query Match 25.6%; Score 278; DB 22; Length 388;
Best Local Similarity 55.6%; Pred. No. 3.2e-18;
Matches 60; Conservative 9; Mismatches 33; Indels 6; Gaps 4;
QY 1 MTNKKAVNPYTNMGKLNPNVGAVVSPDFYAGTVLLCQAN--QEGSSMYSGPSS--VYTS 56
Db 186 MTNKKAVNPYTNMGKLNPNVGAVVSPDFYAGTVLLCQAN--VYTS 56
QY 57 AMPGFPPY-AATAAAYRGALRGRTVYNTFRAAAPPPPIPAYGV 103
Db 246 IIPGFPTAATTAARFAGALRGKRTVYGAVR-AVPEPTIAPAYGV 292

RESULT 14
ID AAM80096 standard; Protein; 286 AA.
XX
AC AAM80096;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 3742.
XX
KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KM tissue growth factor; immunomodulatory; cancer; leukaemia;
KM nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.

PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Aeundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Weinman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR N-PSDB; AAK53229.
XX
XX Claim 20; Page 423; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 286 AA;
Query Match 23.8%; Score 258; DB 22; Length 286;
Best Local Similarity 92.2%; Pred. No. 1.7e-16;
Matches 47; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 MTNKKAVNPYTNMGKLNPNVGAVVSPDFYAGTVLLCQANQEGSSMYSGPSS 51
Db 219 MTNKKAVNPYTNMGKLNPNVGAVVSPDFYAGTVLLCQANQEGSSMYSGPSS 269

RESULT 15
ID ABG29446 standard; Protein; 101 AA.
XX
AC ABG29446;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #29437.
XX
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 N-PSDB; AAS93633.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

PS Claim 20; SEQ ID No 59805; 103pp; English.

XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 101 AA;

Query Match 20.8%; Score 226; DB 22; Length 101;
 Best Local Similarity 50.0%; Pred. No. 5.2e-14;
 Matches 52; Conservative 6; Mismatches 20; Indels 26; Gaps 3;

OY 69 AAAYRGALRGKRTVYNTFRRAAPPPIPAVGWVYQBPVGNKLLGGYAAVRYAQPRT 128
 |||||
 Db 1 AAAYRGALRGKRTVYNTFRRAAPPPIPAVGWVYQDPFYG----- 43

OY 129 PATAAVSDSYG-RVYAADPVYHHTLAPRTYGV-GAMNAFAPLT 170
 :|||
 Db 44 -----ADLYGVMLHTATPSLPPLPPLPTVYVTEFMFSPIS 80

Search completed: March 12, 2003, 22:24:15
 Job time : 40 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2003, 22:23:36 ; Search time 25 Seconds
(without alignments)
780.612 Million cell updates/sec

Title: US-09-809-545A-2

Perfect score: 1086
Sequence: 1 MTNKKAVNPYTNQWKLNPV.....VLSLSQASYOGYNPAPY 203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103	9.5	338	2	C75607 conserved hypothet
2	103	9.5	381	2	KM8 protein rat
3	103	9.5	3124	1	collagen alpha 1(X
4	99	9.1	542	2	hypothetical prote
5	98.5	9.1	655	1	RNA-binding protei
6	98	9.0	316	2	hypothetical prote
7	98	9.0	376	2	segmentation prote
8	97	8.9	811	2	probable transmem
9	96.5	8.9	139	2	extensin precursor
10	96.5	8.9	748	2	hypothetical prote
11	96	8.8	760	2	eye cell developme
12	95	8.7	418	2	hypothetical prote
13	94.5	8.7	469	2	hypothetical prote
14	94	8.7	219	2	adhesin protein -
15	94	8.7	967	2	T20H2.10 protein -
16	93.5	8.6	320	2	chorion protein s3
17	93	8.6	387	2	transcription fact
18	93	8.6	495	1	probable oxidoredu
19	92.5	8.5	348	2	RNA-binding protei
20	92.5	8.5	656	1	hypothetical prote
21	91.5	8.4	325	2	probable PPE prote
22	91.5	8.4	409	2	ELB-55kDa-associat
23	91.5	8.4	856	2	zinc finger RNA bi
24	91.5	8.4	1052	2	eyelid - fruit fly
25	91.5	8.4	2715	2	hypothetical prote
26	91	8.4	164	2	probable secreted
27	91	8.4	748	2	hypothetical prote
28	90.5	8.3	502	2	branched-chain ami
29	89	8.2	318	2	T00626

30	89	8.2	463	2	C70931 probable PPE prote
31	89	8.2	805	2	T25795 hypothetical prote
32	88.5	8.1	390	2	urea/short-chain a
33	88	8.1	2639	2	fibroin - Chinese
34	87.5	8.1	912	2	D72644 hypothetical prote
35	87.5	8.1	1685	2	T43217 RNA polymerase (EC
36	87	8.0	465	2	G02738 FREAC-4 - human
37	87	8.0	482	2	C84330 pantothenate perme
38	86.5	8.0	262	2	A54889 IGE-binding protei
39	86.5	8.0	4302	2	A38971 polycystic kidney
40	86	7.9	568	2	D75627 potassium-transport
41	86	7.9	860	1	EAMS elastin precursor
42	85.5	7.9	345	2	A48462 dense granule prot
43	85.5	7.9	1279	2	T13613 hypothetical prote
44	85	7.8	268	2	A49303 homeotic protein C
45	85	7.8	382	1	SUBSN subtilisin (EC 3.4

ALIGNMENTS

RESULT 1
C75607 conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C/Species: Deinococcus radiodurans
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C/Accession: C75607
R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; Smith, M.; Venter, J.C.; Fraser, C.M.
S./Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A/Reference number: A75250; MUID:20036896; PMID:10567266
A/Accession: C75607
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-338 <WHI>
A/Cross-references: GB:AB001862; GB:AE001825; MID:96460468; PIDN:AF12236.1; PID:96460530
A/Experimental source: strain R1
C/Genetics:
A/Gene: DR0120
A/Map position: 2
C/Superfamily: ymbw protein

Query March 9.5%: Score 103; DB 2; Length 338;
Best Local Similarity 24.9%; Pred. No. 0.3; 78; Indels 36; Gaps 8;
Matches 45; Conservative 22; Mismatches 1; A:SVFDGVLA 161

QY 20 VGAVYSPDFVAGTVLLCOANQSGSSMTSGPSLVYTSMPG-FPYPAATAAAYGAHLR 78
DB 105 LGIGRAPETDRTALALRSRELAADDFPTQIALRFAGETWMP--ACSVFDGVLA 161

QY 79 GRGRTVNTFRPAAPRPPIPAVGVVYOEPIYGNKLLGGYA-AYRAQPTPATPAAYSD 137
DB 162 PRGVYL-----PLTWILGSSLYGAQLAGE--LGVGAFAFYHFSQEDPAPV----- 203

QY 138 SYGRVYADPYPHTLAPR---TYGVGAMNAFAPLTDKTSHADDDVGLVSSLOASTYQ 194
DB 204 -----AVQTYHRHFRFQGLAEPYALIGVNLA-----APTDEARBDLSLSAALSLGILS 253

QY 195 G 195
DB 254 G 254

RESULT 2
JC4647
KM8 protein - rat
N/Alternate names: basic helix-loop-helix protein homolog
C/Species: Rattus norvegicus (Norway rat)
C/Date: 10-May-1996 #sequence_revision 19-Jul-1996 #text_change 20-Jun-2000
C/Accession: JC4647
R/Kume, H.; Maruyama, K.; Tomita, T.; Iwatsubo, T.; Saido, T.C.; Ohta, K.

Biochem. Biophys. Res. Commun. 219, 526-530, 1996
 A>Title: Molecular cloning of a novel basic helix-loop-helix protein from the rat brain.
 A'Reference number: J04647; MUID:96193685; PMID:8605021
 A'Accession: J04647
 A:Molecule type: DNA
 A:Residues: 1-381 <KUM>
 A:Cross-references: DDBJ:D82668; NID:g1166397; PIDN:BA11615.1; PID:g1166398
 A:Experimental source: brain
 C:Comment: This protein is involved in synaptic plasticity, and has a role specific to r-loop-helix domain.
 C:Keywords: brain
 F:122-134/Region: basic
 Query Match 9.5%; Score 103; DB 2; Length 381;
 Best Local Similarity 22.8%; Pred. No. 0.34;
 Matches 39; Conservative 24; Mismatches 76; Indels 32; Gaps 4;
 QY 14 WKLNPFVGVAVSPDFYAGTVLLCOA-----NQESSMSYSPSSL 52
 DB 171 WALSELRLSGKRDLDVSYVOTLCKGLSQPTNLVAGCLQNSRNFTEQGDAAAFTRAR 230
 QY 53 VYTSAPRGFPYPA-----TAAAYRGALRGGR-TVNT-FRAAAPPPYPAVG 101
 DB 231 VARSAMHPRPFCSLAGDSARPPAAGRRGALRTHTGYCAAYETLYAAGGGAAPDYR 290
 QY 102 GVYQBPVYGNKLLQGYAARYAOPTRPATAAYSDSYGRVYADPYHHTL 152
 DB 291 SSEYEGPLSPCLCLNGNFSIAKQSSPDHEKSYHMYHVALPGRAPGHCL 341
 RESULT 3
 A:0020
 collagen alpha 1 (XII) chain precursor - chicken
 N:Alternate names: fibrochimerin
 C:Species: Gallus gallus (Chicken)
 C:Date: 10-Sep-1999 #sequence, revision 10-Sep-1999 #ext change 19-Jan-2001
 C:Accession: A40020; A34485; B34485; A28037; S23814; S22254; S28811
 R:Yamagata, M.; Yamada, K.M.; Yamada, S.S.; Shimomura, T.; Tanaka, H.; Nishida, Y.; Ohara, J. Cell Biol. 115, 209-221, 1991
 A>Title: The complete primary structure of type XII collagen shows a chimeric molecule with house region of type IX collagen, and short collagenous domains with an Arg-Gly-Asp site.
 A'Reference number: A40020; MUID:92011862; PMID:1518137
 A'Accession: A40020
 A:Molecule type: mRNA
 A:Residues: 1-3124 <YAM>
 A:Cross-references: GB:D00824; NID:9222810; PIDN:BA00701.1; PID:9222811
 A>Note: In the authors' translation residues 1216-1219 are shown after residue 1235 and, R:Gordon, M.K.; Gerecke, D.R.; Dublet, B.; van der Rest, M.; Olsen, B.R.
 J. Biol. Chem. 264, 19772-19778, 1989
 A>Title: Type XII collagen. A large multidomain molecule with partial homology to type I A'Reference number: A34485; MUID:90062079; PMID:2584192
 A'Accession: A34485
 A:Molecule type: mRNA
 A:Residues: 2456-2758, 'A', 2760-2802, 'F', 2804-2976, 'F', 2978-3124 <GOR>
 A:Cross-references: EMBL:005137; NID:g211284; PIDN:AAA4635.1; PID:g211285
 A:Accession: B34485
 A:Molecule type: protein
 A:Residues: 2772-2792, 2846-2873 <GOR2>
 R:Gordon, M.K.; Gerecke, D.R.; Olsen, B.R.
 Proc. Natl. Acad. Sci. U.S.A. 84, 6040-6044, 1987
 A>Title: Type XII collagen: distinct extracellular matrix component discovered by cDNA A'Reference number: A28037; MUID:87317590; PMID:3476925
 A'Accession: A28037
 A:Molecule type: mRNA
 A:Residues: 2960-2976, 'F', 2978-3074, 'AG', <GOR3>
 A:Cross-references: EMBL:M17315; NID:g211649; PIDN:AAA46718.1; PID:g211650
 A>Note: This sequence has been revised in reference A34485
 R:Koch, M.; Bernasconi, C.; Chiquet, M.
 Eur. J. Biochem. 207, 847-856, 1992
 A>Title: A major oligomeric fibroblast proteoglycan identified as a novel large form of A'Reference number: S23814; MUID:92362621; PMID:1323460
 A'Accession: S23814
 A:Molecule type: protein

A:Residues: 'X', 1333, 'Q', 1335-1347, 1394-1928, 2504, 'X', 2506, 'X', 2508-2511, 'X', 2513-2517 <R>
 R:Dublet, B.; van der Rest, M.
 J. Biol. Chem. 262, 17724-17727, 1987
 A>Title: Type XII collagen is expressed in embryonic chick tendons. Isolation of pepsin-A'Reference number: S22254; MUID:88087065; PMID:3121603
 A'Accession: S22254
 A:Molecule type: protein
 A:Residues: 2831-2832, 'T', 2834, 'R', 2836-2843, 3002-3014 <DUB>
 R:Trub, J.; Trub, B.
 Biochem. Biophys. Acta 1171, 97-98, 1992
 A>Title: The two splice variants of collagen XII share a common 5' end.
 A'Reference number: S28811; MUID:93042014; PMID:1420368
 A'Accession: S28811
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-24, 1189-1257, 'S', 1259-1263, 'E', 1265-1280 <TRU>
 A:Cross-references: EMBL:X67327
 C:Genetics:
 A:Introns: 2845/3; 2863/3; 2899/3; 2922/1; 2985/1; 3008/1; 3065/1
 C:Superfamily: collagen alpha 1(XII) chain; fibronectin type III repeat homology; von Willebrand; alternative splicing; cell binding; coiled coil; connective tissue; disulfid
 F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-3124/Product: collagen alpha 1(XII) chain #status predicted <MAT>
 F:24-1189-3124/Product: collagen alpha 1(XII) chain short splice form #status predicted
 F:24-114/Domain: IIA #status predicted <IIA>
 F:24-105/Domain: fibronectin type III repeat homology <FN3A>
 F:137-301/Domain: von Willebrand factor type A repeat homology <VWA1>
 F:332-425/Domain: IIB #status predicted <IIB>
 F:332-414/Domain: fibronectin type III repeat homology <FN3B>
 F:437-601/Domain: von Willebrand factor type A repeat homology <VWA2>
 F:629-1178/Domain: IIC #status predicted <IIC>
 F:630-711/Domain: fibronectin type III repeat homology <FN3C>
 F:721-802/Domain: fibronectin type III repeat homology <FN3D>
 F:812-895/Domain: fibronectin type III repeat homology <FN3E>
 F:905-986/Domain: fibronectin type III repeat homology <FN3G>
 F:1095-1076/Domain: fibronectin type III repeat homology <FN3G>
 F:1086-1169/Domain: fibronectin type III repeat homology <FN3H>
 F:1197-1261/Domain: von Willebrand factor type A repeat homology <VWA3>
 F:1384-1465/Domain: IID #status predicted <IID>
 F:1384-1557/Domain: fibronectin type III repeat homology <FN3I>
 F:1474-1557/Domain: fibronectin type III repeat homology <FN3J>
 F:1566-1647/Domain: fibronectin type III repeat homology <FN3K>
 F:1655-1738/Domain: fibronectin type III repeat homology <FN3L>
 F:1756-1838/Domain: fibronectin type III repeat homology <FN3M>
 F:1847-1928/Domain: fibronectin type III repeat homology <FN3N>
 F:1937-2019/Domain: fibronectin type III repeat homology <FN3O>
 F:2028-2110/Domain: fibronectin type III repeat homology <FN3P>
 F:2218-2299/Domain: fibronectin type III repeat homology <FN3Q>
 F:2325-2490/Domain: von Willebrand factor type A repeat homology <VWA4>
 F:2435-2440/Region: cell adhesion #status predicted
 F:2509-2750/Domain: IXP, homologous to NC4 domain of type IX collagen #status predicted
 F:2751-2802/Domain: collagenous COL2 #status predicted <COL2>
 F:2899-2901/Region: cell attachment (R-G-D) motif
 F:2903-2945/Domain: non-collagenous NC2 #status predicted <NC2>
 F:2946-3048/Domain: collagenous COL1 #status predicted <COL1>
 F:3049-3124/Domain: non-collagenous NC1 #status predicted <NC1>
 F:32, 1006, 1032, 1044, 1512, 1767, 2210, 2273, 2532, 2683/Binding site: carbohydrate (Asn) (cova
 F:2780, 2789, 2836, 2842, 2860, 2866, 2869, 3004, 3007/Modified site: hydroxyproline (Pro) #statu

Query Match 9.5%; Score 103; DB 1; Length 3124;
 Best Local Similarity 25.1%; Pred. No. 3.8;
 Matches 42; Conservative 15; Mismatches 46; Indels 64; Gaps 8;
 QY 62 PYPAATAAAYR-----CAHLRGGRV-----YTPFAAAPPPYPAVGCV 104
 DB 2094 PV-KITVAVVDGCGGGLTGNGRTVGLIPONIIYITDEWYFRVSWDPSFVLGYKI 2152
 QY 105 YQBPVGNKLLQ-----GYAARYAOPTRPATAAYSDSYGRVYADPYHHTLAPPTYGVG 161
 DB 2153 VKPVSNGPMFVPGVEVSYTL-----HNLSPSTTYDV- 2186
 QY 162 ANNAFA-----PLTDAKTRSHADVGLVLSLQASLYGCGYNNR 200

A:Reference number: A26636; MUID:87218536; PMID:2884106
 A:Accession: A26636
 A:Molecule type: mRNA
 A:Residues: 1-299 'L', 301-376 <FRA>
 A:Cross-references: GB:X05138; NID:97957; PIDN:CAA28784.1; PID:97958
 C:Genetic8:
 A:Gene: eve (even-skipped)
 A:Cross-references: FlyBase:FBgn0000606
 C:Superfamily: unassigned homeobox proteins; homeobox homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:71-127/Domain: homeobox homology <HOK>

Query Match 9.0%; Score 98; DB 2; Length 376;
 Best Local Similarity 24.3%; Pred. No. 0.88;
 Matches 45; Conservative 8; Mismatches 56; Indels 76; Gaps 7;

QY 22 AYSPDFVGTLLCOANQESSMYSGLVYTSAMPGPYPATTAAYRGALRG 80
 DB 138 AYSDPAPFASITLQAAANSVG-----MPPYAPALAAANAAAA----- 177
 QY 81 GRTVYTFPAAAPPP-----PIPAYGVVYQEPVYGNKLLQGGVAAVRYAQTPTAT 131
 DB 178 --VATNPMAATGMPRGMPQMPTMGPHSGHGHSPYQG-----YKRT----- 220
 QY 132 AAAYSDSYGRVTAADPYHTLAPAPTYGVGAMNAPPLTDAKTRSHADVGLVSLQAS 191
 DB 221 -----FYHTPARPAPHPHAG-----PMHHPHMGSSATGS 251
 QY 192 IYQGG 196
 DB 252 SYSAG 256

RESULT 8
 T36581
 Probable transmembrane protein - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T36581
 R:Oliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrall, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, May 1999
 A:Reference number: 221575
 A:Accession: T36581
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-811 <OLI>
 A:Cross-references: EMBL:AL049826; PIDN:CA842720.1; GSPDB:GN00070; SCODDB:SCH24.16C
 A:Experimental source: strain A3(2)
 C:Genetic8:
 A:Gene: SCODDB:SCH24.16C

Query Match 8.9%; Score 97; DB 2; Length 811;
 Best Local Similarity 27.5%; Pred. No. 2.6;
 Matches 49; Conservative 14; Mismatches 63; Indels 52; Gaps 11;

QY 24 YSPDFVGTLLCOANQESSMYSGLVYTSAMPGPYPATTAAYRGALRG 81
 DB 140 HRPDAFAH---LFRDQGGGHSYDDQAAVAPAPAPG-PYGGAAAPGGY----- 186
 QY 82 KTVYTFPAAAPPPPIPAYGVVYQEPVYGNKLLQGGVAAVRYAQTPTAAAYSDSYGR 141
 DB 187 -----GAPRRPGVYQGHGGRDP-YGNAQTHGQYGG-----SAQYGGG----- 224
 QY 142 VVAADPYHTLAPAP-TYGVGAM--NAPAPLTDAKTRSHADVGLVSLQAS 196
 DB 225 --AATFGQA--APACQIVAPTRPAAEAAPLQEPREPPAAS-----QAAPKKG 270

RESULT 9
 S61885
 extensin precursor (clone 6Pext1.2) - wood tobacco
 C:Species: Nicotiana glauca (wood tobacco)
 C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 19-Jul-2002

C:Accession: S61885
 R:Parmentier, Y.; Durr, A.; Matbach, J.; Hirsinger, C.; Criqui, M.C.; Fleck, J.; Jamet, Plant Mol. Biol. 29, 279-292, 1995
 A:Title: A novel wound-inducible extensin gene is expressed early in newly isolated prot.
 A:Reference number: S61885; MUID:96046747; PMID:7579179
 A:Accession: S61885
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-139 <PAR>
 A:Cross-references: EMBL:X70343
 C:Superfamily: proline-rich protein 3

Query Match 8.9%; Score 96.5; DB 2; Length 139;
 Best Local Similarity 26.1%; Pred. No. 0.37;
 Matches 37; Conservative 12; Mismatches 52; Indels 41; Gaps 6;

QY 34 ILCOANQESSMYSGLVYTSAMPGPYPATTAAYRGALRGRTY 85
 DB 22 LECKANYYSSPPPTTKYVSSPPPYKKSPPPLP-----IY 62
 QY 86 NT-----FRAAPPPPIPAYGVVYQEPVYGNKLLQGGVAAVRYAQTPTAAAYSDSYG 140
 DB 63 RSPPPVYKYSPPPIPKYSSPPPPVYKSP-----PPVYKYSPPPPVYKSP 117
 QY 141 RYVYADP-----YHTLAPAPTY 158
 DB 118 PVKSPPPPYHYTSPPPPHY 139

RESULT 10
 T04011
 Hypothetical protein TSL19.200 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999
 R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Meyer, K.F.X
 submitted to the Protein Sequence Database, March 1999
 A:Reference number: Z15184
 A:Accession: T04011
 A:Molecule type: DNA
 A:Residues: 1-748 <BEV>
 A:Cross-references: EMBL:AL049481
 A:Experimental source: cultivar Columbia; BAC clone TSL19
 C:Genetic8:
 A:Map position: 4
 A:Introns: 67/2; 209/3; 271/3; 305/3; 329/3; 367/3; 691/1
 A>Note: TSL19.200

Query Match 8.9%; Score 96.5; DB 2; Length 748;
 Best Local Similarity 26.9%; Pred. No. 2.6;
 Matches 46; Conservative 16; Mismatches 70; Indels 39; Gaps 7;

QY 18 PVGVYSPDFVGTLLCOANQESSMYSGLVYTSAMPGPYPATTAAYR 73
 DB 584 PYQATPAPAPSYGSTNMAPQQQGYTSSDGFVQOQTPSYSSAPSDAYNNGTOTPATG 643
 QY 74 GAHLRGRTVYNTF-RAAAPPPPIPAYGVV-----YQEPVYGNKLLQGGVAAV 123
 DB 644 PAYQOQSVQPASTYDQGAQAAAGCGVAPRGTYTPRSQAYSGQ-----AAYS 697
 QY 124 VAOPT-----PATAAVSDSYGRVYAADPYHTLAPAPTYGVGAMNAPFA 167
 DB 698 QAAPQTQYEQQAPATQAA-----VYATAP-----GTAPVKTQSPOSAXYA 736

RESULT 11
 A45174
 eye cell development gene eye protein, splice form 1 - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C:Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998
 R:Bonini, N.M.; Leiserson, W.M.; Benzer, S.
 Cell 72, 379-395, 1993

A>Title: The eyes absent gene: genetic control of cell survival and differentiation in *C. elegans*
 A:Reference number: A45174; MUID:93161413; PMID:8431945
 A:Accession: A45174
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1760 <BON>
 A:Cross-references: GB:U08501; NID:g157975; PID:g157976
 A:Experimental source: adult head
 A:Note: Sequence extracted from NCBI backbone (NCBI:124850)
 C:Genetics:
 A:Gene: FLYBase:cli; eya; eyes absent
 A:Cross-references: FlyBase:FBgn0000320

Query Match 8.8%; Score 96; DB 2; Length 760;
 Best Local Similarity 23.7%; Pred. No. 2.9;
 Matches 42; Conservative 24; Mismatches 61; Indels 50; Gaps 6;

QY 42 GSSMY-----SGSSSLVYTSAMPGFPPATTAAGALRGRTVNTFRAAA 92
 DB 223 GSNLYGSSASNPIDGAVAVNSSAV-----AAAAAAYDGGK-----DYIYNSMOOYT 272
 QY 93 PPPPIYAGVYVQEPYVGNKLLQGYAAYRVAOPTPATATA----- 134
 DB 273 PPPPYSGYGTPTAAATARQAKMEPGAAAAAAYLTPTSYAASGNNNSQLYSSFYAGYNNF 332
 QY 135 -----YSDSYGRVYAADPYHHTLAPAP-----TYGVG-AMNAFAPLTDKTRSHA 178
 DB 333 GQODYGVYVNGYNYSPANYSPYAVSSPSSASHGHGFVAASNSNSESPTDTHS 389

RESULT 12
 T15142
 hypothetical protein T28F2.6 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jan-2000

C:Accession: T15142
 R:Madson, C.; Fironick, B.
 submitted to the EMBL Data Library, April 1997
 A:Description: The sequence of *C. elegans* cosmid T28F2.
 A:Reference number: Z18300
 A:Accession: T15142
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-418 <MAD>
 A:Cross-references: EMBL:AF00198; NID:G2047345; PID:G2047346; PIDN:AB53052.1; GSPDB:GN
 C:Experimental source: strain Bristol N2; clone T28F2
 C:Genetics:
 A:Gene: CESP:T28F2.6
 A:Map position: 1
 A:Insertions: 49/3
 C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 8.7%; Score 95; DB 2; Length 418;
 Best Local Similarity 30.7%; Pred. No. 1.8;
 Matches 42; Conservative 7; Mismatches 60; Indels 28; Gaps 5;

QY 40 QEGSMYSMPESLVYTSAMP-----FPYPATTAAGALRGGR 82
 DB 272 KDGAGPDGPAGCTDAGPDDAAYCCPPRTTGAGAYRPGDAAAAPAGYDGGAGA 331
 QY 83 TVVNTFRAAAPPPPIYAGVYVQEPYVGNKLLQGYAAYRVAOPTPATTAAYSDSYGRV 142
 DB 332 APEAAPAAAAAPAPAPAAAAA--APAGG---YQGG-AAAGAAAPPPAPAAAAAEPAPAA 385
 QY 143 YAADPYHHTLAPAPTYG 159
 DB 386 PAAAP-----PPAPAG 397

RESULT 13
 T20047
 hypothetical protein C49A1.4 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T20047
 R:Mathews, L.
 submitted to the EMBL Data Library, December 1996
 A:Reference number: Z19217
 A:Accession: T20047
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-469 <MIL>
 A:Cross-references: EMBL:Z83221; PIDN:CAB05707.1; GSPDB:GN00019; CESP:C49A1.4
 A:Experimental source: clone C49A1
 C:Genetics:
 A:Gene: CESP:C49A1.4
 A:Map position: 1
 A:Insertions: 37/1; 98/1; 140/3; 249/3; 318/1; 441/3

Query Match 8.7%; Score 94.5; DB 2; Length 469;
 Best Local Similarity 23.6%; Pred. No. 2.2;
 Matches 51; Conservative 22; Mismatches 80; Indels 63; Gaps 9;

QY 18 PVVGAVYSPDFYAGTVLLCOANOEGSSMY---SGPSSLVYTSAMP----- 59
 DB 25 PTVDLAISEAYGSTSTSTLTSSVTQYNSYPOYAMTSAHPANYVQGVNTANLACT 84
 QY 60 -GPPYPATTAAGALRGRTVNTFRAAAPPPPIYAGVYVQEPYVGNKLLQGG 118
 DB 85 TAFPYSLTT--PSYGSY-----PVDYTSAAAA-----YQNPYYTN--LRGG 122
 QY 119 YAAARYAOPTPATAAASDSYGRVYAADPYHHTLAPAPTYGV--GAMNAFA-----PLTDA 172
 DB 123 TAAPYNPPLNATTAAYAAVAAVSSVLTGDAVNLGTSDSGSGVPSVTISFSLKEKPKVSK 182
 QY 173 KTRSHA-----DDVGLVLSLQASI 192
 DB 183 KKTGSCSPDEYARVFMIDIDIAVISRNLYASV 218

RESULT 14
 S42674
 adhesive protein - bifurcate mussel (fragments)
 N:Alternate names: polyphenolic protein
 C:Species: *Septifer bifurcatus* (bifurcate mussel)
 C>Date: 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 07-May-1999
 C:Accession: S42674
 R:Repecki, L.M.; Chiu, S.S.; Waite, J.H.; Lavin, M.F.
 Mol. Marine Biol. Biotechnol. 1, 78-88, 1991
 A>Title: Molecular diversity of marine glues: polyphenolic proteins from five mussel spec
 A:Reference number: S42671; MUID:93364572; PMID:1845474
 A:Accession: S42674
 A:Molecule type: protein
 A:Residues: 1-13;14-36;37-39;40-52;53-65;66-78;79-81;82-86;87-94;95-101;102-109;110-117;1
 A:Note: Ser-19 was also found

Query Match 8.7%; Score 94; DB 2; Length 219;
 Best Local Similarity 28.7%; Pred. No. 1;
 Matches 39; Conservative 14; Mismatches 47; Indels 36; Gaps 9;

QY 46 YSGPSSLVYTSAMPGFPPATTAAGALRGGR-----RTVYNT--FRAAP 93
 DB 64 YKYPTR--YTYGTGYRPAKAPKTSYGTYGYSYGTGYKSSYGTGYKAPAK 121
 QY 94 PPPPIYAGVYVQEPV-----YGNKLLQGYAAYRVAOPTPATTAAYSDSYGRVYAA 145
 DB 122 PPSKPSYGTGYKAPKTYTTPSSYGT-----GYKA-----PTKYTTK--PSSYGTGYKA 169
 QY 146 DPYHHTLAPAPTYGVG 161
 DB 170 -PTKYSTKPS-SYGTG 183

RESULT 15
 H86334
 T20H2.10 protein - *Arabidopsis thaliana*

C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001
 C;Accession: H86334
 R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luross, J.S.; Malci, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A;Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.
 A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A;Reference number: A86141; MUID:21016719; PMID:11130712
 A;Accession: H86334
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-967 <STO>
 A;Cross-references: GB:AE005172; NID:98778986; PIDN:AAF79901.1; GSPDB:GN00141
 C;Genetics:
 A;Map position: 1

Query March 8.7%; Score 94; DB 2; Length 967;
 Best Local Similarity 25.4%; Pred. No. 5.6;
 Matches 50; Conservative 20; Mismatches 77; Indels 50; Gaps 10;
 QY 18 PVVGAIVSPPDF--YAGTVLLCOANOGSSMYSGPSLSVYTSAMPGFYPATATAAAYRGA 75
 Db 54 PPTGCGSADYENYQNTYFYQNSE---HYPSAPSFSPSQPPSPATSL----- 103
 QY 76 HLRGGRIVYNTFRAAAPPP-----PIPAYG---VYQEP---VYGNKLQGGYAY 122
 Db 104 ----NPNYSSTFNOPPPPTIHPQPLSSYSGSFDSSTAPYQCPSTQCHMYSPYDOHOTSGY 158
 QY 123 RYAOPTPATAAAYSDSYGVVADPYHHTLAPRTYGVGAMNAFAPLTDKTRSHADVYG 182
 Db 159 SSA-PPSSAPAPNPN-----PAPYSSSLYAPPYSSGSS--IPPSYKPSVKFPDQSG 209
 QY 183 LVLSLQASIVYGGYNR 199
 Db 210 -----YDGYNR 215

Search completed: March 13, 2003, 00:54:42
 Job time : 27 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2003, 21:21:40 ; Search time 16 Seconds
(without alignments)

526.231 Million cell updates/sec

Title: US-09-809-545A-2

Perfect score: 1086
Sequence: 1 MTNRKAVNPYTNMGKLNPPV.....VLSSLQASIVQGGYRFPAPY 203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1020.5	94.0	396	1 A2BP_MOUSE	Q9J343 mus musculu
2	1009.5	93.0	397	1 A2BP_HUMAN	Q9HWD1 homo sapien
3	307.5	28.3	391	1 RBM9_HUMAN	O43251 homo sapien
4	116	10.7	3063	1 CALC_HUMAN	O99715 homo sapien
5	115	10.6	3119	1 CALC_MOUSE	O60847 mus musculu
6	103	9.5	3124	1 CALC_CHICK	P13944 gallus gall
7	98.5	9.1	376	1 FXE1_HUMAN	O00338 homo sapien
8	98.5	9.1	655	1 EMS_MOUSE	O61545 mus musculu
9	98	9.0	376	1 HMEV_DROME	O63252 rattus norv
10	97	8.9	497	1 BRN1_RAT	O05201 drosophila
11	96	8.8	766	1 EYA_DROME	O62424 mus musculu
12	94	8.7	386	1 HKAD_MOUSE	P31271 homo sapien
13	94	8.7	388	1 HXAD_MOUSE	P17110 ceratilis c
14	93.5	8.6	320	1 CH36_CERCA	P33693 gallus gall
15	93	8.6	387	1 GAT6_CHICK	P31361 mus musculu
16	93	8.6	495	1 BRN1_MOUSE	O01844 homo sapien
17	92.5	8.5	656	1 EMS_HUMAN	P20264 homo sapien
18	92	8.5	500	1 BRN1_HUMAN	Q92754 homo sapien
19	90.5	8.3	450	1 AP2C_HUMAN	P56645 homo sapien
20	90.5	8.3	1210	1 Y102_MYCTU	O53951 mycobacteri
21	89	8.2	463	1 Y102_MYCTU	P97334 mus musculu
22	87.5	8.1	362	1 NK2C_MOUSE	O16676 homo sapien
23	87	8.0	465	1 FXD1_HUMAN	P08699 rattus norv
24	86.5	8.0	261	1 LEG3_RAT	Q9ZRN7 desinococcu
25	86	7.9	568	1 ATVA_DEIRA	P54320 mus musculu
26	86	7.9	860	1 ELIS_MOUSE	Q27002 toxoplasma
27	85.5	7.9	345	1 GR44_TOXGO	P18111 mus musculu
28	85	7.8	268	1 CDX1_MOUSE	P00763 bacillus am
29	85	7.8	382	1 SUBT_BACAM	P16376 drosophila
30	85	7.8	746	1 TUP2_DROME	P20469 panacea ana
31	85	7.8	1322	1 ICEA_PAVAN	P81445 alcaigenes
32	84.5	7.8	330	1 NIR_ALCXX	P07518 bacillus pu
33	84	7.7	275	1 SUBT_BACPU	

34	84	7.7	876	1 BGL1_SACFI	P22506 saccharomyc
35	83.5	7.7	466	1 ANX7_HUMAN	P20073 homo sapien
36	83.5	7.7	520	1 LAC1_TRAVI	O99044 trameetes vi
37	83.5	7.7	741	1 TLE4_RAT	O07141 rattus norv
38	83.5	7.7	766	1 TLE4_HUMAN	O04727 homo sapien
39	83.5	7.7	766	1 TLE4_MOUSE	O62441 mus musculu
40	83	7.6	450	1 FAX8_HUMAN	O06710 homo sapien
41	83	7.6	463	1 ANX7_MOUSE	O07076 mus musculu
42	83	7.6	621	1 VP40_HSVBC	P54817 bovine herp
43	82.5	7.6	520	1 LAC1_CORHI	O02497 coriolus hi
44	82.5	7.6	538	1 SNPH_HUMAN	O15079 homo sapien
45	82.5	7.6	604	1 GLAS_DROME	P13360 drosophila

ALIGNMENTS

RESULT 1
A2BP_MOUSE STANDARD; PRT; 396 AA.
ID A2BP_MOUSE
AC Q9J343; Q9JTB5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ataxin 2-binding protein.
GN A2BP1 OR A2BP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Oada N., Kusuda J., Tanuma R., Ito A., Hirata M., Sugano S.,
RA Hashimoto K.;
RT "Isolation of full-length cDNA clones from mouse brain cDNA library
made by oligo-capping method."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 21-396 FROM N.A.
RX MEDLINE=2075619; PubMed=10814712;
RA Shihbata H., Huynh D.P., Pulst S.-M.;
RT "A novel protein with RNA-binding motifs interacts with ataxin-2."
RL Hum. Mol. Genet. 9:1303-1313(2000).
CC -1- SUBMIT: BINDS TO THE C-TERMINAL OF ATAXIN-2.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB041596; BAA95079.1; -;
DR EMBL: AF107204; AAF78292.1; -;
DR HSPB; P11940.1; IGV;
DR MGD; MGI:1926224; A2bp1.
DR InterPro; IPR005054; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PSS0102; RRM; 1.
DR PROSITE; PSS0030; RRM_RNP_1; 1.
DR RNA-binding.
KW DOMAIN 116 192 RNA-BINDING (RRM).
FT CONFLICT 60 G -> D (IN REF. 2).
FT CONFLICT 128 R -> RFR (IN REF. 2).
FT CONFLICT 369 G -> S (IN REF. 2).
SQ SEQUENCE 396 AA; 42590 MM; 9A0C59C5E687F39F CRC64;
Query Match 94.0%; Score 1020.5; DB 1; Length 396;
Best Local Similarity 94.6%; Pred. No. 1.8e-74;

Matches 192; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

QY 1 MTNKKAVNPYTNMGKLNPNVGAIVSPDFYAGTVLLCOANOESSSMVSGSSLVYTSAMP 60
 DB 195 MTNKKAVNPYTNMGKLNPNVGAIVSPDFYAGTVLLCOANOESSSMVSGSSLVYTSAMP 254

QY 61 PPYPATAAAAYRGALHGRGRVTYNTFRPAAAPPPPIPAYGGVVOEFPYGNKLLGGYA 120
 DB 255 PPYPATAAAAYRGALHGRGRVTYNTFRPAAAPPPPIPAYGGVVOEFPYGNKLLGGYA 313

QY 121 AYRYAOPPTATAAASDSYGRVYAADPYHHTLAPATYGVGMNAPFLTDKTRSHAD 180
 DB 314 AYRYAOPPTATAAASDSYGRVYAADPYHHTLAPATYGVGMNAPFLTDKTRSHAD 373

QY 181 VGLVLSLQASLYOGGYNRFAPY 203
 DB 374 VGLVLSLQASLYOGGYNRFAPY 396

RESULT 2
 A2BP_HUMAN STANDARD; PRT; 397 AA.
 AC Q9NWB1; Q9NS20;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ataxin-2-binding protein.
 GN A2BP1 OR A2BP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RA Isegaki T., Ota T., Hayashi K., Sugiyama T., Otauki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shitatori A., Sudo H., Sugawara M.,
 RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Konodo H., Sugawara M.,
 RA Takehashi M., Chiba Y., Ishida S., Murakawa K., Ono W., Takiguchi S.,
 RA Takehara S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuo Y.,
 RA Ninomiya K., Iwayanagi T.;
 RT "NEDO human cDNA sequencing project";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP MEDLINE=20275619; PubMed=10814712;
 RX MEDLINE=20275619; PubMed=10814712;
 RA Shibata H., Huynh D.P., Pulst S.-M.;
 RT "A novel protein with RNA-binding motifs interacts with ataxin-2";
 RL Hum. Mol. Genet. 9:1303-1313(2000).
 CC -1- SUBUNIT: BINDS TO THE C-TERMINAL OF ATAXIN-2.
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MUSCLE AND BRAIN.
 CC -1- SIMILARITY: CONTRAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; AK001027; BAA91472.1; -;
 DR EMBL; AF107203; AAF78291.1; -;
 DR HSSP; P11940; ICVJ.
 DR MIM; 605104; -;
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; Rrm; 1.
 DR SMART; SMO0360; RRM; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP_1; 1.
 KM RNA-binding.
 FT DOMAIN 117 193 RNA-BINDING (RRM).
 FT CONFLICT 92 92 A -> T (IN REF. 2).

SQ SEQUENCE 397 AA; 42754 MW; E3E9060B68C79880 CRC64;

Query Match 93.0%; Score 1009.5; DB 1; Length 397;
 Best Local Similarity 93.6%; Pred. No. 1,4e-73;
 Matches 190; Conservative 4; Mismatches 8; Indels 1; Gaps 1;

QY 1 MTNKKAVNPYTNMGKLNPNVGAIVSPDFYAGTVLLCOANOESSSMVSGSSLVYTSAMP 60
 DB 196 MTNKKAVNPYTNMGKLNPNVGAIVSPDFYAGTVLLCOANOESSSMVSGSSLVYTSAMP 255

QY 61 PPYPATAAAAYRGALHGRGRVTYNTFRPAAAPPPPIPAYGGVVOEFPYGNKLLGGYA 120
 DB 256 PPYPATAAAAYRGALHGRGRVTYNTFRPAAAPPPPIPAYGGVVOEFPYGNKLLGGYA 314

QY 121 AYRYAOPPTATAAASDSYGRVYAADPYHHTLAPATYGVGMNAPFLTDKTRSHAD 180
 DB 315 AYRYAOPPTATAAASDSYGRVYAADPYHHTLAPATYGVGMNAPFLTDKTRSHAD 374

QY 181 VGLVLSLQASLYOGGYNRFAPY 203
 DB 375 VGLVLSLQASLYOGGYNRFAPY 397

RESULT 3
 RBM9_HUMAN STANDARD; PRT; 391 AA.
 AC Q43251; Q9UGW4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Putative RNA-binding protein 9 (RNA binding motif protein 9).
 GN RBM9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RA Whiteley M.;
 RT Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP MEDLINE=20275619; PubMed=10814712;
 RX MEDLINE=20275619; PubMed=10814712;
 RA Whiteley M.;
 RT "A novel protein with RNA-binding motifs interacts with ataxin-2";
 RL Hum. Mol. Genet. 9:1303-1313(2000).
 CC -1- SUBUNIT: BINDS TO THE C-TERMINAL OF ATAXIN-2.
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MUSCLE AND BRAIN.
 CC -1- SIMILARITY: CONTRAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; AL009266; CAAL5842.1; ALT_SEQ.
 DR EMBL; AL049748; CAB63054.1; -;
 DR HSSP; P11940; ICVJ.
 DR Genes; HGNC:9906; RBM9.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; Rrm; 1.
 DR SMART; SMO0360; RRM; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP_1; 1.
 KM RNA-binding.
 FT DOMAIN 112 188 RNA-BINDING (RRM).
 FT SEQUENCE 391 AA; 42284 MW; ED2B76C03D16A7B CRC64;

Query Match 28.3%; Score 307.5; DB 1; Length 391;
 Best Local Similarity 43.0%; Pred. No. 1.3e-17;
 Matches 77; Conservative 13; Mismatches 56; Indels 31; Gaps 7;

QY 1 MTNKKAVNPYTNMGKLNPNVGAIVSPDFYAGTVLLCOAN--OEGSSMVGSSSLVYTSAMP 56

```

Db 191 MTNKKWTPYANGKLSVGVAVGVPELHYAASSFOADVSLGNDAVPLSGRGQINTYIPL 250
Qy 57 AMGEPVP-AATAAAAYGAHLRGRTVYNTFRAAAPPIPGVYGVDEPYGNLL 115
Db 251 IIPGEPPTAATTAATTAAGAHLRGRGRTVYGAVR-AVPTAIIPALPGVYDGGFYGLADLY 309
Qy 116 QGGAAYRV---AQPTATTAAYSDSYGRVYAADPHYHTLAPATYGVGMNAPAPLT 170
Db 310 IESANCRSNRVMDQPT-----DMHSLLLQPPQ---LLQPLQPLT 347

RESULT 4
CAIC_HUMAN STANDARD; PRT: 3063 AA.
AC 099715; 099716.
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(XII) chain precursor.
GN COL12A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (LONG AND SHORT FORM), AND PARTIAL SEQUENCE.
RX MEDLINE=97288521; PubMed=9143499;
RA Gerecke D.R., Olson P.F., Koch W., Knoll J.H.M., Taylor R.,
RA Hudson D.L., Champilaud M.-F., Olsen B.R., Burgess R.E.;
RT "Complete primary structure of two splice variants of collagen XII,
RT and assignment of alpha 1(XII) collagen (COL12A1), alpha 1(XI)
RT collagen (COL9A1), and alpha 1(XIX) collagen (COL19A1) to human
RT chromosome 6q12-q13."
RL Genomics 41:236-242(1997).
CC -1- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-
CC CONTAINING FIBRILS. THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE
CC SURFACE OF THE FIBRILS, AND THE COL2 AND NC3 DOMAINS MAY BE
CC LOCALIZED IN THE PERIFIBRILLAR MATRIX (BY SIMILARITY).
CC -1- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 KDa OF
CC NONTRIPLE-HELICAL SEQUENCES.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE FINAL TISSUE
CC FORM OF COLLAGEN XII MAY CONTAIN HOMOTRIMERS OF EITHER THE LONGER
CC OR THE SHORTER ISOFORM OR ANY COMBINATION OF LONG AND SHORT
CC ISOFORM CHAINS.
CC -1- TISSUE SPECIFICITY: FOUND IN COLLAGEN I-CONTAINING TISSUES: BOTH
CC SHORT AND LONG ISOFORMS APPEAR IN AMNION, CHORION, SKELETAL
CC MUSCLE, SMALL INTESTINE, AND IN CELL CULTURE OF DERMAL
CC FIBROBLASTS, KERATINOCYTES, AND ENDOTHELIAL CELLS. ONLY THE SHORT
CC ISOFORM IS FOUND IN LONG, PLACENTA, KIDNEY, AND A SQUAMOUS CELL
CC CARCINOMA CELL LINE.
CC -1- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT
CC EACH END (BY SIMILARITY).
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY
CC SIMILARITY).
CC -1- PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY). ONLY THE
CC LONG VARIANT IS A PROTEOGLYCAN.
CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC INTERRUPTED HELICES (FACIT) FAMILY.
CC -1- SIMILARITY: CONTAINS 4 VMPA DOMAINS.
CC -1- SIMILARITY: CONTAINS 18 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; U73778; AAC51244.1; -

```

```

DR EMBL; U73779; AAD40483.1; -.
DR HSSP; P02751; 1TTF.
DR Genew; HGNC:2188; COL12A1.
DR MIM; 120320; -.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003962; FNIII_repeat.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR02035; VWF_A.
DR Pfam; PF00041; Fn3; 18.
DR Pfam; PF00092; Wnt4; 4.
DR Pfam; PF01391; Collagen; 4.
DR Pfam; PF02210; TSPN; 1.
DR PRINTS; PR00014; FNTYPEIII.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00060; FN3; 17.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00327; VWF; 4.
DR PROSITE; PS50234; VMPA; 4.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.
FT SIGNAL 1 24
FT CHAIN 25 3063 COLLAGEN ALPHA 1(XII) CHAIN.
FT DOMAIN 25 114 FIBRONECTIN TYPE-III 1.
FT DOMAIN 140 316 VMPA 1.
FT DOMAIN 333 426 FIBRONECTIN TYPE-III 2.
FT DOMAIN 440 616 VMPA 2.
FT DOMAIN 630 721 FIBRONECTIN TYPE-III 3.
FT DOMAIN 722 812 FIBRONECTIN TYPE-III 4.
FT DOMAIN 813 903 FIBRONECTIN TYPE-III 5.
FT DOMAIN 904 998 FIBRONECTIN TYPE-III 6.
FT DOMAIN 999 1085 FIBRONECTIN TYPE-III 7.
FT DOMAIN 1086 1178 FIBRONECTIN TYPE-III 8.
FT DOMAIN 1199 1371 VMPA 3.
FT DOMAIN 1384 1473 FIBRONECTIN TYPE-III 9.
FT DOMAIN 1474 1564 FIBRONECTIN TYPE-III 10.
FT DOMAIN 1565 1652 FIBRONECTIN TYPE-III 11.
FT DOMAIN 1654 1751 FIBRONECTIN TYPE-III 12.
FT DOMAIN 1752 1842 FIBRONECTIN TYPE-III 13.
FT DOMAIN 1843 1932 FIBRONECTIN TYPE-III 14.
FT DOMAIN 1933 2023 FIBRONECTIN TYPE-III 15.
FT DOMAIN 2024 2114 FIBRONECTIN TYPE-III 16.
FT DOMAIN 2115 2202 FIBRONECTIN TYPE-III 17.
FT DOMAIN 2203 2291 FIBRONECTIN TYPE-III 18.
FT DOMAIN 2323 2496 VMPA 4.
FT DOMAIN 2451 2746 NONHELICAL REGION (NC3)
FT DOMAIN 2747 2898 TRIPLE-HELICAL REGION (COL2)
FT DOMAIN 2899 2941 WITH 1 IMPERFECTION.
FT DOMAIN 2942 3044 NONHELICAL REGION (NC2).
FT DOMAIN 3045 3063 TRIPLE-HELICAL REGION (COL1)
FT BINDING 798 889 WITH 2 IMPERFECTIONS.
FT BINDING 889 981 NONHELICAL REGION (NC1).
FT SITE 862 981 TO CHONDROITIN SULFATE (POTENTIAL).
FT SITE 2779 2781 TO CHONDROITIN SULFATE (POTENTIAL).
FT SITE 2895 2897 CELL ATTACHMENT SITE (POTENTIAL).
FT MOD_RES 2895 2944 CELL ATTACHMENT SITE (POTENTIAL).
FT MOD_RES 2944 2944 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2947 2947 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2950 2950 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2959 2959 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2965 2965 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2968 2968 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2971 2971 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2983 2983 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3000 3000 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3003 3003 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3014 3014 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3023 3023 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3026 3026 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3029 3029 HYDROXYLATION (BY SIMILARITY).
FT CARBOHYD 700 N-LINKED (GLCNAC. .) (POTENTIAL).

```

```

FT CARBOHYD 1763 1763 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2206 2206 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2528 2528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2679 2679 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 25 1188 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 3063 AA; 333189 MW; 75FEA78FA8548293 CRC64;

Query Match 10.7%; Score 116; DB 1; Length 3063;
Best Local Similarity 25.1%; Pred. No. 0.21;
Matches 45; Conservative 16; Mismatches 54; Indels 64; Gaps 8;

Oy 50 SLVVTSMAMPGRPPPAATAAAAYR---GAHLGRGRTV-----YNTFRAAA 92
Db 2078 NNVIQLPPLDPFPY-KITYIAVEDDGGHLTGNGRTVGLPPONHISDEWYRFRVSW 2136
Oy 93 PPPPIPVAGGVVYOEPRVNVKLLQ---GGVAAVRYAQTPRATAAASDYSGRVYAADPYH 149
Db 2137 DDPSPVULCYKIVYKRVGNSNEMEAHVGMTSTYL----- 2171
Oy 150 HTLAPAPTYGVGAMNAFA-----PLTDKTRSHADVDGLVLSLQASIQGGYNRP 200
Db 2172 HMLNSTYTDV---NVYAQYDSGLSVPLTDQGT-----LYLVNTDKTYQIGWDTF 2220

RESULT 5
CALC MOUSE STANDARD; PRT; 3119 AA.
ID 060847; P70322;
AC 15-JUN-1998 (Rel. 36, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(XII) chain precursor.
GN COL12A1.
OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RP (1)
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS XI1A-1 AND
RP XI1B-1).
RC STRAIN=Swiss Webster, and C57BL/6J; TISSUE=Skin;
RX MEDLINE=96170761; PubMed=8610106;
RA Boelme K., Li Y., Oh P.S., Olsen B.R.;
RA "Primary structure of the long and short splice variants of mouse
RT collagen XII and their tissue-specific expression during embryonic
RT development.";
RL Dev. Dyn. 204:432-445 (1995).
RN [2]
RN PARTIAL SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS XI1A-2
RN AND XI1B-2).
RP STRAIN=C57BL/6J; TISSUE=Skin fibroblast;
RX MEDLINE=99348349; PubMed=10419532;
RA Kania B.R., Reichenberger E., Baur S.T., Karimbux N.Y., Taylor R.W.,
RA Olsen B.R., Nishimura I.;
RA "Structural variation of type XII collagen at its carboxyl-terminal
RT NCI domain generated by tissue-specific alternative splicing.";
RL J. Biol. Chem. 274:22053-22059 (1999).
CC -1- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-
CC CONTAINING FIBRILS, THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE
CC SURFACE OF THE FIBRILS, AND THE COL2 AND NC3 DOMAINS MAY BE
CC LOCALIZED IN THE PERIFIBRILLAR MATRIX (BY SIMILARITY).
CC -1- SUNUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 kDa OF
CC NONTRIPE-HELICAL SEQUENCES (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS: XI1A-1 (SHOWN HERE), XI1A-
CC 2/ERNA, XI1B-1 AND XI1B-2, ARE PRODUCED BY ALTERNATIVE SPLICING.
CC THE FINAL TISSUE FORM OF COLLAGEN XII MAY CONTAIN HOMOTRIMERS OR
CC ANY COMBINATION OF THE VARIOUS ISOFORMS.
CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN TENDONS, PERICHONDRUM,
CC SKIN, CORNEA, SCLERA, BLOOD VESSELS, AND PERIOSTEUM.
CC -1- DEVELOPMENTAL STAGE: THE LONG NC3 XI1A ISOFORMS ARE PREDOMINANT AT
CC EARLY STAGES (ED7 AND 11); AT LATER STAGES OF DEVELOPMENT (ED15
CC AND 17) THE SHORT NC3 XI1B FORMS BECOME THE MAJOR FORMS. AS THE
CC SHORT NC3 FORMS BECOME THE MAJOR PRODUCT, THE LONG SPLICED VARIANT

```

```

CC CONTINUES TO BE EXPRESSED IN SEVERAL TISSUES, EVEN AFTER BIRTH.
CC THE LONG NC1 ISOFORMS, XI1A-1 AND XI1B-1, PEAK IN 15-DAY OLD
CC EMBRYOS AND DECREASE IN 17-DAY OLD ONES. THE EXPRESSION OF THE
CC SHORT NC1 FORM XI1B-2 REMAINS CONSTANT THROUGHOUT LATE STAGES OF
CC EMBRYONIC DEVELOPMENT (ED15 AND ED17).
CC -1- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT
CC EACH END (BY SIMILARITY).
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY
CC SIMILARITY).
CC -1- PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY). ONLY
CC ISOFORM XI1A-2 IS A PROTEOGLYCAN.
CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC INTERRUPTED HELICES (FACIT) FAMILY.
CC -1- SIMILARITY: CONTAINS 4 WMFA DOMAINS.
CC -1- SIMILARITY: CONTAINS 18 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; U25652; AAA9719.1; ALT_SEQ.
CC EMBL; U57095; AAB07047.1; -.
CC HSRP; P02751; IFNA.
CC MGD; MGI:88448; Col12a1.
CC InterPro; IPR000087; Col12a1.
CC InterPro; IPR003961; FN_III.
CC InterPro; IPR003962; FN_III_repeat.
CC InterPro; IPR003129; TSPN.
CC InterPro; IPR002035; WMFA.
CC Pfam; PF00041; Fn3; 18.
CC Pfam; PF00092; Waa; 4.
CC Pfam; PF01391; Collagen; 4.
CC Pfam; PF02210; TSPN; 1.
CC PRINTS; PR00014; FNTYPEIII.
CC PRINTS; PR00453; WMFADOMAIN.
CC SMART; SMO0060; FN3; 16.
CC SMART; SMO0210; TSPN; 1.
CC SMART; SMO0327; WAA; 4.
CC PROSITE; PS50234; WMFA; 4.
CC KX Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.
CC KW
CC SIGNAL
CC CHAIN
CC 25 3119
CC FT DOMAIN 25 114 FIBRONECTIN TYPE-III 1.
CC FT DOMAIN 140 316 WMFA 1.
CC FT DOMAIN 332 423 FIBRONECTIN TYPE-III 2.
CC FT DOMAIN 444 620 WMFA 2.
CC FT DOMAIN 634 725 FIBRONECTIN TYPE-III 3.
CC FT DOMAIN 726 816 FIBRONECTIN TYPE-III 4.
CC FT DOMAIN 817 907 FIBRONECTIN TYPE-III 5.
CC FT DOMAIN 908 1002 FIBRONECTIN TYPE-III 6.
CC FT DOMAIN 1003 1089 FIBRONECTIN TYPE-III 7.
CC FT DOMAIN 1090 1182 FIBRONECTIN TYPE-III 8.
CC FT DOMAIN 1203 1375 WMFA 3.
CC FT DOMAIN 1388 1474 FIBRONECTIN TYPE-III 9.
CC FT DOMAIN 1480 1568 FIBRONECTIN TYPE-III 10.
CC FT DOMAIN 1569 1652 FIBRONECTIN TYPE-III 11.
CC FT DOMAIN 1659 1757 FIBRONECTIN TYPE-III 12.
CC FT DOMAIN 1758 1848 FIBRONECTIN TYPE-III 13.
CC FT DOMAIN 1849 1938 FIBRONECTIN TYPE-III 14.
CC FT DOMAIN 1939 2029 FIBRONECTIN TYPE-III 15.
CC FT DOMAIN 2030 2120 FIBRONECTIN TYPE-III 16.
CC FT DOMAIN 2121 2208 FIBRONECTIN TYPE-III 17.
CC FT DOMAIN 2209 2297 FIBRONECTIN TYPE-III 18.
CC FT DOMAIN 2456 2751 NONHELICAL REGION (NC3).
CC FT DOMAIN 2752 2899 TRIPLE-HELICAL REGION (COL2)
CC WITH 1 IMPERFECTION.

```

```

FT DOMAIN 2900 2942 NONHELICAL REGION (NC2).
FT DOMAIN 2943 3045 TRIPLE-HELICAL REGION (COL1)
FT 3046 3119 WITH 2 IMPERFECTIONS.
FT BINDING 893 893 NONHELICAL REGION (NC1).
FT BINDING 893 893 TO CHONDROITIN SULFATE (POTENTIAL).
FT BINDING 893 893 TO CHONDROITIN SULFATE (POTENTIAL).
FT SITE 866 868 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2784 2786 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2896 2898 CELL ATTACHMENT SITE (POTENTIAL).
FT MOD_RES 2945 2945 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2945 2945 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2951 2951 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2951 2951 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2960 2960 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2966 2966 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2969 2969 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2972 2972 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 2984 2984 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3001 3001 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3004 3004 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3015 3015 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3024 3024 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3027 3027 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 3030 3030 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 869 872 POLY-THR.
FT CARBOHYD 704 704 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1769 1769 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2212 2212 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2533 2533 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2684 2684 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 25 1190 MISSING (IN ISOFORM X1IB-1 AND ISOFORM
FT VARSPLIC 3062 3064 X1IB-2).
FT VARSPLIC 3064 3064 EPI -> GSG (IN ISOFORM X1IA-2 AND ISOFORM
FT VARSPLIC 3119 3119 X1IB-2).
FT VARSPLIC 3119 3119 MISSING (IN ISOFORM X1IA-2 AND ISOFORM
SQ SEQUENCE 3119 AA; 340239 NM; 981F99C86A3251 CRC64;
Query Match 10.6%; Score 115; DB 1; Length 3119;
Best Local Similarity 24.4%; Pred. No. 0.26; Mismatches 59; Indels 58; Gaps 7;
Matches 43; Conservative 16;
QY 50 SGLVYTSAMPGEPPYPAATAAAYR---GAHLGRGRV-----YNTFRAAA 92
DB 2084 NNVLILQLPQPTPY-KIVIVAIYEDGDGHLGNRTYGLAPQNIHIFDEMYEYCFRFSW 2142
QY 93 PPPPIPAYGVVYQEPVYGNKLLQ---GQYAAARYAQPPATPAAAYSSYGRVYAADPYH 149
DB 2143 DPSPSPVLYGYKIVYKPGVSNPEMFAFVGEVTSYTL----- 2177
QY 150 HTLAPAPYVGAMNAF-----APLTDKTRSHADVDGLVLSLQASYOGGVNRF 200
DB 2178 HNLNPSTTVDSVYAQYDGLSVPLTDGTT-----LYLNVTDLKYQVGWLTFF 2226

```

```

RA Yamagata M., Yamada K.M., Yamada S.S., Shinomura T., Tanaka H.,
RA Nishida Y., Obata M., Kinata K.;
RT "The complete primary structure of type XII collagen shows a chimeric
RT molecule with reiterated fibronectin type III motifs, von Willebrand
RT factor A motifs, a domain homologous to a noncollagenous region of
RT type IX collagen, and short collagenous domains with an Arg-Gly-Asp
RT site.";
RT J. Cell Biol. 115:209-221(1991).
RN [2]
RN SEQUENCE OF 2456-3124 FROM N.A., AND SEQUENCE OF 2772-2794 AND
RP 2446-2873.
RX MEDLINE=90062079; Pubmed=2584192;
RA Gordon M.K., Gerecke D.R., Dublet B., van der Reest M., Olsen B.R.;
RT "Type XII collagen. A large multidomain molecule with partial
RT homology to type IX collagen.";
RL J. Biol. Chem. 264:19772-19778(1989).
RN [3]
RN SEQUENCE OF 2960-3076 FROM N.A.
RX MEDLINE=87317590; Pubmed=3476925;
RA Gordon M.K., Gerecke D.R., Olsen B.R.;
RT "Type XII collagen: distinct extracellular matrix component
RT discovered by cDNA cloning.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:6040-6044(1987).
RN [4]
RN SEQUENCE OF 1-1283 FROM N.A. (SHORT FORM), AND ALTERNATIVE SPLICING.
RP TISSUE=Embryo;
RX MEDLINE=93042014; Pubmed=1420368;
RA Trueb J., Trueb B.;
RT "The two splice variants of collagen XII share a common 5' end.";
RL Biochim. Biophys. Acta 1171:97-98(1992).
RN [5]
RN ALTERNATIVE SPLICING.
RX MEDLINE=95370352; Pubmed=7642694;
RA Koch M., Bohmann B., Mathieson M., Hagios C., Trueb B., Chiquet M.;
RT "Large and small splice variants of collagen XII: differential
RT expression and ligand binding.";
RL J. Cell Biol. 130:1005-1014(1995).
RN [6]
RN FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-
CC CONTAINING FIBRILS. THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE
CC SURFACE OF THE FIBRILS, AND THE COL2 AND NC3 DOMAINS MAY BE
CC LOCALIZED IN THE PERIFIBRILLAR MATRIX.
CC [7]
CC SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 KDA OF
CC NONTRIPE-HELICAL SEQUENCES.
CC [8]
CC ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE FINAL TISSUE
CC FORM OF COLLAGEN XII MAY CONTAIN HOMOTRIMERS OF EITHER THE LONGER
CC OR THE SHORTER ISOFORM OR ANY COMBINATION OF LONG AND SHORT
CC ISOFORM CHAINS. ONLY THE LONG VARIANT IS A PROTEOLYTICALLY THE LARGE
CC ISOFORM HAS MORE RESTRICTED EXPRESSION IN EMBRYONIC TISSUE THAN
CC THE SMALL.
CC [9]
CC TISSUE SPECIFICITY: TYPE XII COLLAGEN IS PRESENT IN TENDONS,
CC LIGAMENTS, PERICHONDRUM, AND PERIOSTEUM, ALL DENSE CONNECTIVE
CC TISSUES CONTAINING TYPE I COLLAGEN.
CC [10]
CC DOMAIN: THIS SEQUENCE DEFINES FIVE DISTINCT DOMAINS. TWO TRIPLE-
CC HELICAL DOMAINS (COL1 AND COL2) AND THREE NONTRIPE-HELICAL
CC DOMAINS (NC1, NC2, AND NC3).
CC [11]
CC PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT
CC EACH END.
CC [12]
CC PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC [13]
CC PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY).
CC [14]
CC SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC INTERRUPTED HELICES (FACIT) FAMILY.
CC [15]
CC SIMILARITY: CONTAINS 4 WMPA DOMAINS.
CC [16]
CC SIMILARITY: CONTAINS 18 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC [17]
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

CC -----
 DR EMBL; D00824; BAA0701.1; -
 DR EMBL; X61024; CAA43358.1; -
 DR EMBL; M17375; AAA48718.1; -
 DR EMBL; J05137; AAA48635.1; -
 DR EMBL; X67327; CAA47744.1; -
 DR PIR; A28037; A28037.
 DR PIR; A34485; A34485.
 DR HSSP; P20701; ILPA.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR003962; FN_III_repeat.
 DR InterPro; IPR003129; TSPN.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF00041; fn3; 17.
 DR Pfam; PF00092; vwa; 4.
 DR Pfam; PF01391; Collagen; 4.
 DR Pfam; PF02210; TSPN; 1.
 DR PRINTS; PR00014; FNYPEI11.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00060; FN3; 17.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00327; VWA; 4.
 DR PROSITE; PS50234; VWF_A; 4.
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.
 FT SIGNAL 1 24
 FT CHAIN 25 3124
 FT DOMAIN 25 114
 FT DOMAIN 139 311
 FT DOMAIN 332 425
 FT DOMAIN 439 615
 FT DOMAIN 629 720
 FT DOMAIN 721 811
 FT DOMAIN 812 904
 FT DOMAIN 905 998
 FT DOMAIN 999 1085
 FT DOMAIN 1086 1178
 FT DOMAIN 1179 1371
 FT DOMAIN 1384 1473
 FT DOMAIN 1474 1565
 FT DOMAIN 1566 1654
 FT DOMAIN 1655 1755
 FT DOMAIN 1756 1846
 FT DOMAIN 1847 1936
 FT DOMAIN 1937 2027
 FT DOMAIN 2028 2118
 FT DOMAIN 2119 2206
 FT DOMAIN 2207 2295
 FT DOMAIN 2296 2387
 FT DOMAIN 2388 2475
 FT DOMAIN 2476 2560
 FT DOMAIN 2561 2650
 FT DOMAIN 2651 2740
 FT DOMAIN 2741 2830
 FT DOMAIN 2831 2920
 FT DOMAIN 2921 2992
 FT DOMAIN 2993 3086
 FT DOMAIN 3087 3176
 FT DOMAIN 3177 3266
 FT DOMAIN 3267 3356
 FT DOMAIN 3357 3446
 FT DOMAIN 3447 3536
 FT DOMAIN 3537 3626
 FT DOMAIN 3627 3716
 FT DOMAIN 3717 3806
 FT DOMAIN 3807 3896
 FT DOMAIN 3897 3986
 FT DOMAIN 3987 4076
 FT DOMAIN 4077 4166
 FT DOMAIN 4167 4256
 FT DOMAIN 4257 4346
 FT DOMAIN 4347 4436
 FT DOMAIN 4437 4526
 FT DOMAIN 4527 4616
 FT DOMAIN 4617 4706
 FT DOMAIN 4707 4796
 FT DOMAIN 4797 4886
 FT DOMAIN 4887 4976
 FT DOMAIN 4977 5066
 FT DOMAIN 5067 5156
 FT DOMAIN 5157 5246
 FT DOMAIN 5247 5336
 FT DOMAIN 5337 5426
 FT DOMAIN 5427 5516
 FT DOMAIN 5517 5606
 FT DOMAIN 5607 5696
 FT DOMAIN 5697 5786
 FT DOMAIN 5787 5876
 FT DOMAIN 5877 5966
 FT DOMAIN 5967 6056
 FT DOMAIN 6057 6146
 FT DOMAIN 6147 6236
 FT DOMAIN 6237 6326
 FT DOMAIN 6327 6416
 FT DOMAIN 6417 6506
 FT DOMAIN 6507 6596
 FT DOMAIN 6597 6686
 FT DOMAIN 6687 6776
 FT DOMAIN 6777 6866
 FT DOMAIN 6867 6956
 FT DOMAIN 6957 7046
 FT DOMAIN 7047 7136
 FT DOMAIN 7137 7226
 FT DOMAIN 7227 7316
 FT DOMAIN 7317 7406
 FT DOMAIN 7407 7496
 FT DOMAIN 7497 7586
 FT DOMAIN 7587 7676
 FT DOMAIN 7677 7766
 FT DOMAIN 7767 7856
 FT DOMAIN 7857 7946
 FT DOMAIN 7947 8036
 FT DOMAIN 8037 8126
 FT DOMAIN 8127 8216
 FT DOMAIN 8217 8306
 FT DOMAIN 8307 8396
 FT DOMAIN 8397 8486
 FT DOMAIN 8487 8576
 FT DOMAIN 8577 8666
 FT DOMAIN 8667 8756
 FT DOMAIN 8757 8846
 FT DOMAIN 8847 8936
 FT DOMAIN 8937 9026
 FT DOMAIN 9027 9116
 FT DOMAIN 9117 9206
 FT DOMAIN 9207 9296
 FT DOMAIN 9297 9386
 FT DOMAIN 9387 9476
 FT DOMAIN 9477 9566
 FT DOMAIN 9567 9656
 FT DOMAIN 9657 9746
 FT DOMAIN 9747 9836
 FT DOMAIN 9837 9926
 FT DOMAIN 9927 10016
 FT DOMAIN 10017 10106
 FT DOMAIN 10107 10196
 FT DOMAIN 10197 10286
 FT DOMAIN 10287 10376
 FT DOMAIN 10377 10466
 FT DOMAIN 10467 10556
 FT DOMAIN 10557 10646
 FT DOMAIN 10647 10736
 FT DOMAIN 10737 10826
 FT DOMAIN 10827 10916
 FT DOMAIN 10917 11006
 FT DOMAIN 11007 11096
 FT DOMAIN 11097 11186
 FT DOMAIN 11187 11276
 FT DOMAIN 11277 11366
 FT DOMAIN 11367 11456
 FT DOMAIN 11457 11546
 FT DOMAIN 11547 11636
 FT DOMAIN 11637 11726
 FT DOMAIN 11727 11816
 FT DOMAIN 11817 11906
 FT DOMAIN 11907 12000

FT CARBOHYD 2683 2683 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VANSPLIC 25 1188 MISSING (IN SHORT ISIFORM).
 FT CONFLICT 1258 1258 T -> S (IN REF. 4).
 FT CONFLICT 1264 1264 D -> E (IN REF. 4).
 FT CONFLICT 2759 2759 P -> A (IN REF. 2).
 FT CONFLICT 2803 2803 L -> F (IN REF. 2).
 FT CONFLICT 2977 2977 V -> F (IN REF. 2).
 FT CONFLICT 3075 3076 OP -> AG (IN REF. 3).
 SQ SEQUENCE 3124 AA; 340578 MW; 094285AF7F346CF CRC64;
 Query Match 9.5%; Score 103; DB 1; Length 3124;
 Best Local Similarity 25.1%; Pred. No. 2.3;
 Matches 42; Conservative 15; Mismatches 46; Indels 64; Gaps 8;
 QY 62 PYPAATAAAYR---GAHLRGGRIV-----YTFPAAAPPPPIPAYGVV 104
 DB 2094 PY-KITVAAYVEDGCGGLTGNGRTVGLPPONITYIDEMWYRFVSDPSPVILGYKI 2152
 QY 105 YOEPIYGNKLDL---CGYAAHYAQPPTATAAAYSDSGRYVAADPHHTLAPAPTYGVG 161
 DB 2153 VYKPVGSNEPMVEFVGEVTSYTL-----HNLSPTTYDV- 2186
 QY 162 ANNAFA-----PLTDKTRSHADVGLVLSLQASLYGGYNRF 200
 DB 2187 --NVTAQDSCMSITLDGTT-----LYNVTDLTYYKGMPTF 2224
 RESULT 7
 FXEL HUMAN STANDARD; PRT; 376 AA.
 AC 000158; 075765;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Forkhead box protein E1 (Forkhead-related protein FKHL15) (Thyroid
 DE transcription factor 2) (TFP-2).
 GN FOXE1 OR FKHL15 OR TTF2 OR TTF2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=keratinocytes;
 RX MEDLINE=97312696; PubMed=9169137;
 RA Chadwick B.P.; Obermayr F.; Frischaut A.-M.;
 RT "FKHL15, a new human member of the forkhead gene family located on
 RT chromosome 9q22.4";
 RL Genomics 41:390-396(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Macchia P.E.; Mattei M.-G.; Lapi P.; Fenzi G.; Di Lauro R.;
 RT "Cloning, chromosomal localization and identification of two
 RT polymorphisms in the human thyroid transcription factor 2 gene
 RT (TTF2).";
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP VARIANT BAMFORTH-LAZARUS SYNDROME VAL-65.
 RX MEDLINE=98361170; PubMed=9697705;
 RA Clifton-Bligh R.J.; Wentworth J.M.; Heinz P.; Criap M.S.; John R.;
 RA Lazarus J.H.; Ludgate M.; Charterjee V.K.;
 RT "Mutation of the gene encoding human TTF-2 associated with thyroid
 RT agenesis, cleft palate and choanal atresia.";
 RL Nat. Genet. 19:399-401(1998).
 CC -!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR. COULD BE INVOLVED IN
 CC THYROID GLAND ORGANOGENESIS.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- TISSUE SPECIFICITY: DETECTED IN ADULT BRAIN, PLACENTA, LUNG,
 CC LIVER, SKELETAL MUSCLE, KIDNEY, PANCREAS, HEART, COLON, SMALL
 CC INTESTINE TESTIS AND THYMUS. EXPRESSION WAS STRONGEST IN HEART AND
 CC PANCREAS.
 CC -!- DISEASE: DEFECTS IN FOXE1 ARE THE CAUSE OF BAMFORTH-LAZARUS
 CC SYNDROME. A DISEASE ASSOCIATED WITH THYROID AGENESIS, CLEFT PALATE

CC AND CHOMAL ATRESTA.
CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U98995; AAC51294.1; -;
DR EMBL: Y13386; CAA73816.1; -;
DR HSSP: Q63245; ZHFH.
DR TRANSFAC: T02782; -;
DR Genew: HGNC:3806; FOXE1.
DR MIM: 602617; -;
DR MIM: 241850; -;
DR InterPro: IPR001766; TF_Fork_head.
DR Pfam: PF00250; Fork_head; 1.
DR PRINTS: PR00053; FORKHEAD.
DR ProDom: PD000425; TF_Fork_head; 1.
DR SMART: SM00339; FH; 1.
DR PROSITE: PS00657; FORK_HEAD_1; 1.
DR PROSITE: PS00658; FORK_HEAD_2; 1.
DR PROSITE: PS50039; FORK_HEAD_3; 1.
KM DNA-binding; Nuclear protein; Transcription regulation;
KM Disease mutation.
FT DOMAIN 7 12 POLY-PRO.
FT DNA_BIND 52 143 FORK-HEAD.
FT DOMAIN 164 182 POLY-ALA.
FT VARIANT 65 65 A->V (IN BAMFORTH-LAZARUS SYNDROME).
FT /FTID=VAR.008857.
FT CONFLICT 178 182 MISSING (IN REF. 2).
SQ SEQUENCE 376 AA; 38289 MM; 766534A73231572C C6C64;

Query Match 9.1%; Score 98.5; DB 1; Length 376;
Best Local Similarity 28.3%; Pred. No. 0.55;
Matches 45; Conservative 13; Mismatches 68; Indels 33; Gaps 6;

QY 12 NGWLTNPVGVAVSPDFYAGTVLLCQANQEGSSMYSGSSLYVTSAMGFPYPAATAAA 71
DB 125 NYMALDPVAEDM---FSSGSFLRRKRKPSDL-----STYPAYMMDAAAAAA 170
QY 72 YRGALHRCGRGVNTFFAAAPPPPIPAYGVVQEPYGVNKLQGVAAVRYQPTPAT 131
DB 171 AAAAAAIAAIFPGAVPAARPP---YPGAIV-----AGYAPPSLAPPPVY 215
QY 132 AAAYSDSYGRVYA--ADPYHTTLAPAPTYGVGANNAPA 167
DB 216 YPAASPGPCRVFGLVPERPLSPGLPAPS-GPGSCAFA 253

RESULT 8
ID EMS_MOUSE STANDARD; PRT; 655 AA.
AC 061545;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE RNA-binding protein Ems.
GN Ems OR EMSG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=95130099; PubMed=7829909;
RA Ploungastel B., Mattei M.-G., Thomas G., Delattre O.,
RT "Cloning and chromosome localization of the mouse Ews gene.",
RL Genomics 23:278-281(1994).

CC -1- FUNCTION: MIGHT FUNCTION AS A REPRESSOR (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -1- SIMILARITY: CONTAINS 1 RANBP2-TYPE ZINC FINGER.
CC -1- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X79233; CAA55815.1; -;
DR MGI: 99960; Eweb.
DR InterPro: IPR000504; RNA_rec_mot.
DR InterPro: IPR001876; Znf_RanBP.
DR Pfam: PF00076; rrm; 1.
DR Pfam: PF00641; zf-RanBP; 1.
DR SMART: SM00360; RRM; 1.
DR SMART: SM00547; Znf_RBP2; 1.
DR PROSITE: PS50066; IQ; FALSE_NEG.
DR PROSITE: PS50102; RRM; 1.
DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
DR PROSITE: PS01358; ZF_RANBP2_1; 1.
DR PROSITE: PS50199; ZF_RANBP2_2; 1.
KM Nuclear protein; zinc-finger; zinc; RNA-binding; Metal-binding;
KM Calmodulin-binding; Repeat; Methylation; Phosphorylation;
KM Transcription regulation; Repressor.
FT DOMAIN 1 285 EAD (GLN/PRO/THR-RICH).
FT DOMAIN 256 285 IQ.
FT ZN_FING 360 446 RNA-BINDING (RRM).
FT 517 548 RANBP2-TYPE.
FT 8 285 31 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 8 16 1.
FT REPEAT 17 27 2.
FT REPEAT 28 34 3.
FT REPEAT 35 42 4.
FT REPEAT 43 50 5.
FT REPEAT 51 59 6.
FT REPEAT 60 68 7.
FT REPEAT 69 75 8.
FT REPEAT 76 84 9.
FT REPEAT 85 91 10.
FT REPEAT 92 110 11.
FT REPEAT 111 116 12.
FT REPEAT 117 125 13.
FT REPEAT 126 156 14.
FT REPEAT 157 163 15.
FT REPEAT 164 170 16.
FT REPEAT 171 177 17.
FT REPEAT 178 188 18.
FT REPEAT 189 193 19.
FT REPEAT 194 201 20.
FT REPEAT 202 206 21.
FT REPEAT 207 212 22.
FT REPEAT 213 218 23.
FT REPEAT 219 224 24.
FT REPEAT 225 230 25.
FT REPEAT 231 238 26.
FT REPEAT 239 245 27.
FT REPEAT 246 252 28.
FT REPEAT 253 259 29.
FT REPEAT 260 276 30.
FT REPEAT 277 285 31.
FT DOMAIN 300 339 ARG/GLY/PRO-RICH.
FT DOMAIN 453 512 ARG/GLY/PRO-RICH.
FT DOMAIN 558 639 ARG/GLY/PRO-RICH.
FT MOD_RES 300 300 METHYLATION (DI-).
FT MOD_RES 302 302 METHYLATION (DI-).
FT MOD_RES 304 304 METHYLATION (DI-).
FT (BY SIMILARITY).
FT (BY SIMILARITY).
FT (BY SIMILARITY).

```

FT MOD_RES 309 309 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 314 314 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 317 317 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 321 321 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 454 454 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 463 463 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 470 470 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 489 489 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 493 493 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 499 499 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 502 502 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 505 505 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 562 562 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 564 564 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 571 571 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 574 574 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 580 580 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 588 588 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 591 591 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 595 595 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 599 599 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 602 602 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 606 606 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 614 614 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 632 632 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 635 635 METHYLATION (DI-) (BY SIMILARITY)
FT MOD_RES 266 266 PHOSPHORYLATION (BY PKC) (BY SIMILARITY)
SQ SEQUENCE 655 AA; 68418 MW; 50735EB54247D69 CRC64;

```

```

Query Match 9.1%; Score 98.5; DB 1; Length 655;
Best Local Similarity 28.5%; Pred. No. 1;
Matches 49; Conservative 11; Mismatches 71; Indels 41; Gaps 9;

```

```

QY 29 VAGTLLTQCA-NOEGSSWYSGPSLVTYAMGEPFPAATAAAYGALHRCGRVTNT 87
DB 29 YAGTT---OAYGQOSGYTGGPTDVSYTQHTTATTCGTATVST-----GQPTGYST 79
QY 88 FRAAAPPPPIPAVGVGVYQVYVGNKLLQGYAAVYVYVYVYVYVYVYVYVYVYVYV 147
DB 80 -----FPAFQ-----AYGQPVGY-----GTGAYVSTTATVTTTQA-----SYVAQS 116
QY 148 YHHTLAPATYGVGAMNATAPLTDATRSHADVGLVLSLQASTYQGYNR 199
DB 117 AYGTGPAYVPTGYO-----QPTATAPTRPDGNGKPAETSPQSS--TGGYNQ 160

```

RESULT 9

```

HMEV_DROME STANDARD; PRT; 376 AA.
AC P06602; P07667; Q9V5B6;
DT 01-JAN-1988 (Rel. 06, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Segmentation protein even-skipped.
GN EVE OR CG3238.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_taxid=7227;
OK NCBI_taxid=7227;
RN RP SEQUENCE FROM N.A.
RX MEDLINE=87051744; PubMed=2877745;
RA McDonald P.M., Ingham P., Struhl G.;
RT "Isolation, structure, and expression of even-skipped: a second pair-
RL rule gene of Drosophila containing a homeo box.";
RL Cell 47:721-734 (1986).
RN RP SEQUENCE FROM N.A.
RX MEDLINE=87218536; PubMed=2884106;
RA Prach M., Hoey T., Rushlow C., Doyle H., Levine M.;
RT "Characterization and localization of the even-skipped protein of

```

```

RT Drosophila.";
RL EMBO J. 6:749-759 (1987).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Gale R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer G., Chapple M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bonbrant D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brodtier P.,
RA Burkova D., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dudin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jaitli M., Kalush F., Karpen G.H., Ke Z., Kemstun J.A., Ketchum K.A.,
RA Jaisli B., Kalush F., Karpen G.H., Ke Z., Kemstun J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Liaw P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Paclab J.M.,
RA Palazzolo M., Plittman G.S., Pan S., Pollard J., Puri V., Shen M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shee H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkac R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach U.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Zheng H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [4]
RP SEQUENCE OF 1-58 FROM N.A.
RC STRAIN=OR-RC, WA-F, ZMS6, AF-S, and FL-S;
RX MEDLINE=96036621; PubMed=8524036;
RA Ludwig M.Z., Kreitman M.;
RT "Evolutionary dynamics of the enhancer region of even-skipped in
RL Drosophila.";
RL Mol. Biol. Evol. 12:1002-1011 (1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 70-189.
RX MEDLINE=96134926; PubMed=8557047;
RA Hirsch J.A., Aggarwal A.K.;
RT "Structure of the even-skipped homeodomain complexed to AT-rich DNA:
RT new perspectives on homeodomain specificity.";
RL EMBO J. 14:6280-6291 (1995).
CC -I- FUNCTION: MAY PLAY A ROLE IN DETERMINING NEURONAL IDENTITY. MAY BE
CC DIRECTLY INVOLVED IN SPECIFYING IDENTITY OF INDIVIDUAL NEURONS.
CC PAIR-RULE PROTEIN REQUIRED FOR SEGMENTATION; INVOLVED IN
CC TRANSFORMING THE BROAD, SPATIAL, APERIODIC EXPRESSION PATTERNS OF
CC THE GAP GENES INTO A SYSTEM OF PRECISE PERIODIC EXPRESSION
CC PATTERNS OF THE PAIR-RULE AND SEGMENTARY POLARITY GENES.
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC -I- SIMILARITY: BELONGS TO THE EVEN-SKIPPED HOMEOBOX FAMILY.
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```


Db 303 NSHDPHSDPTSDPDLQPAKOFKORIKLQFTQADVGLALGTLYGNVF 352

RESULT 11

EYA_DROME STANDARD: PRT: 766 AA.

ID EYA_DROME 005201; Q9VMCL1; 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Developmental protein eyes absent (Protein Clift).

GN CLIF OR EYA OR CG9554.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;

OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

NCBI_TaxID=7227;

CC [1]

RN SEQUENCE FROM N.A.

RP MEDLINE=93161413; PubMed=8431945;

RA Bonini N.M., Leiserson W.M., Benzer S.;

RT "The eyes absent gene: genetic control of cell survival and

RT differentiation in the developing Drosophila eye.";

RT Cell 72:379-395(1993).

FT [2]

RN SEQUENCE FROM N.A.

RP STRAIN=Berkeley;

RC MEDLINE=20196006; PubMed=10731132;

RX Adame M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktoglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck H., Butler E., Center A., Chandra I.,

RA Buttle K.C., Busan D.A., Butler H., Cadelis E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo J., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downe M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrelira S., Fleischmann W.,

RA Fostel R.A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasseer K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hooten D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jaitli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodaira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milhine N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svendsen R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RT Science 287:2185-2195(2000).

CC -1- FUNCTION: REQUIRED AT AN EARLY STAGE IN DEVELOPMENT OF THE

CC D. MELANOGASTER COMPOUND EYE.

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE

CC PRODUCED BY ALTERNATIVE SPLICING.

CC -1- SIMILARITY: BELONGS TO THE EYA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.ebi.ac.uk/announce/>

CC or send an email to license@ebi.ac.uk).

CC -----

CC DR EMBL, L08501; AAA28723.1; -

CC DR EMBL, L08502; AAA28110.1; -

CC DR EMBL, AE003614; AAF52400.1; -

CC DR FLYBase; FBgn0000320; c1i.

CC DR InterPro: IPR001454; Hlgase/hydrase.

CC Pfam; PF00702; Hydrolase, 1.

CC Developmental protein; Nuclear protein; Alternative splicing.

CC KW DOMAIN 46 57 POLY-GLN.

CC FT DOMAIN 60 68 POLY-GLN.

CC FT DOMAIN 92 108 POLY-GLY.

CC FT DOMAIN 253 260 POLY-ALA.

CC FT DOMAIN 268 271 POLY-TYR.

CC FT DOMAIN 305 311 POLY-ALA.

CC FT DOMAIN 428 434 POLY-ALA.

CC FT VARSPLIC 1 23 MVTLMPPYVAAPRCGLIDKMIETP -> MLYVPCYQNFSTL

CC FT DYI (IN ISOFORM 2).

CC SQ SEQUENCE 766 AA; 80656 MW; 3581C26AB1811E74 CRC64;

Query Match 8.8%; Score 96; DB 1; Length 766;

Best Local Similarity 23.7%; Pred. No. 1.9;

Matches 42; Conservative 24; Mismatches 61; Indels 50; Gaps 6;

QY 42 GSSM-----SGPSLYVTSMPGPYPATRAAAGAHLRGRGTYNFRRAA 92

Db 229 GSNLGGSSASNPDLGVAANVSSAV-----AAAAVYDCKH---DYIYNSMOQYT 278

QY 93 PPPPIPAYGVVYQPPVYGNKLLGCGYAAARYAQPPTATAA----- 134

Db 279 PPPFYSGVCTPPAAATTAARQAAMEGCAAAAAAAYLTPTAAAGNNNSQLYSPVAGYNNF 338

QY 135 -----YSDYSGRYVADPPYHHTLAPAP---TYGCG-AMNAPAPLTDATRTSHA 178

Db 339 GQDYGYYNEQYGNVYSPANYPYAVSPSSSSASHGCHFWAASSNLSSEPTDHS 395

RESULT 12

HXAD_MOUSE STANDARD: PRT: 386 AA.

ID HXAD_MOUSE 062424;

AC 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Homeobox protein Hox-A13 (Hox-1.10).

GN HOXA13 OR HOX-1.10.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI_TaxID=10090;

RN [1]

RN SEQUENCE FROM N.A.

RP MEDLINE=96259555; PubMed=8673126;

RA Mortlock D.P., Post L.C., Innis J.W.;

RT "The molecular basis of hypodactyly (Hd): a deletion in Hoxa 13 leads

RT to arrest of digital arch formation.";

RT Nat. Genet. 13:284-289(1996).

CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF

CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH

CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS (BY

CC SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- DISEASE: DEFECTS IN HOXA13 ARE THE CAUSE OF HYPODACTYLY (HD), A

CC CONDITION CHARACTERIZED BY PROFOUND DEFICIENCY OF DIGITAL ARCH

CC STRUCTURES.

CC -1- SIMILARITY: BELONGS TO THE HXD-B HOMEBOX FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U59322; AAB03322.1; -;
DR HSSP; P14653; 1B72.
DR TRANSFAC; T03337; -;
DR MGD; MGI:96173; Hoxa13.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PSS00027; HOMEBOX_1; 1.
DR PROSITE; PSS0071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation.
FT DNA BIND 38 51 HOMEBOX.
FT DOMAIN 52 57 POLY-ALA.
FT DOMAIN 62 66 POLY-GLY.
FT DOMAIN 73 84 POLY-ALA.
FT DOMAIN 101 104 POLY-ALA.
FT DOMAIN 116 133 POLY-ALA.
FT DOMAIN 198 205 POLY-ALA.
SQ SEQUENCE 386 AA; 39566 MW; 2B01DCC9B1951324 CRC64;

Query Match 8.7%; Score 94; DB 1; Length 386;
Best Local Similarity 23.9%; Pred. No. 1.3;
Matches 56; Conservative 15; Mismatches 91; Indels 72; Gaps 9;

QY 21 GAVYSPDFYACTVLLCOANGSSMYSGPSL-----VYTSAMPGEPPYPAATAAAAYRG 76
DB 67 GGNSTVAAAAAANQCRNLMAHPALFAGAAAYSSAPGAPPSAAAAA 125
QY 77 LRGGRFTYNTFRAAAPPPPIPAYGVVYQ-----EPYGNKLLGGYAAVRY-- 124
DB 126 -----AAAAAASSSGGPGPAPAGAAKQCSPCSAQAQSSSGPALPFGYFGSGYPCA 182
QY 125 -----AQTPTAAAYSDS-----YGRVYAADPYNHT 151
DB 183 RMGHPNAIKSCAPASAAAFADKWDGTGPAAEFFSSAKKEPAFYHOGYAAAGPYHHN 241
QY 152 LAAPPTY-----GVGAMNAFAPLTDKTRSHADVDGLVLSIQASLYQGYN 198
DB 242 -QPVPGYLDMPVVPGLGGPG-----ESRHEPLGLPMESYQPMALPNGWN 284

RESULT 13
HXAD_HUMAN STANDARD; PRT; 388 AA.
ID AC P31271; O43371;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein Hox-A13 (Hox-13).
GN HOXA13 OR HOX13.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP MEDLINE=97172976; PubMed=9020844;
RA Mortlock D.P., Innis J.W.;
RT "Mutation of HOXA13 in hand-foot-genital syndrome.";
RL Nat. Genet. 15:179-180(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Bradshaw H., Hinds K., Keppler D.;

RL Submitted (FE8-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 322-387 FROM N.A.
RX MEDLINE=9009876; PubMed=2574852;
RA Acampora D., D'Esposito M., Faiella A., Pannese M., Migliaccio E.,
RA Morelli F., Stornaiuolo A., Nigro V., Simeone A., Boncinelli E.;
RT "The human HOX gene family.";
RL Nucleic Acids Res. 17:10385-10402(1989).
CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE ABD-B HOMEBOX FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U82827; AAC50993.1; -;
DR EMBL; AC004080; -; NOT_ANNOTATED_CDS.
DR PIR; S14932; S14932.
DR HSSP; P14653; 1B72.
DR TRANSFAC; T03321; -;
DR Genem; HGNC:5102; HOXA13.
DR MIM; 142959; -;
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PSS00027; HOMEBOX_1; 1.
DR PROSITE; PSS0071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation.
FT DNA BIND 322 381 HOMEBOX.
FT DOMAIN 38 53 POLY-ALA.
FT DOMAIN 62 66 POLY-ALA.
FT DOMAIN 73 84 POLY-ALA.
FT DOMAIN 116 133 POLY-ALA.
FT DOMAIN 145 150 POLY-ALA.
FT DOMAIN 200 207 POLY-ALA.
FT CONFLICT 146 146 A -> G (IN REF. 2).
FT CONFLICT 187 187 P -> H (IN REF. 2).
FT CONFLICT 195 195 P -> A (IN REF. 2).
FT CONFLICT 198 198 P -> A (IN REF. 2).
SQ SEQUENCE 388 AA; 39752 MW; 6CD9C9A5616C2FF6 CRC64;

Query Match 8.7%; Score 94; DB 1; Length 388;
Best Local Similarity 23.6%; Pred. No. 1.3;
Matches 56; Conservative 18; Mismatches 87; Indels 76; Gaps 10;

QY 21 GAVYSPDFYACTVLLCOANGSSMYSGPSL-----VYTSAMPGEPPYPAATAAAAYRG 74
DB 67 GGNSTVAAAAAANQCRNLMAHPALFAGAAAYSSA-PGEAPPSAAAAA 124
QY 75 LRGGRFTYNTFRAAAPPPPIPAYGVVYQ-----EPYGNKLLGGYAAVRY-- 124
DB 125 A-----AAAAAASSSGGPGPAGPAAAEAAKQCSPCSAQAQSSSGPALPFGYFGSGYYP 180
QY 125 -----AQTPTAAAYSDS-----YGRVYAADPY 148
DB 181 CARMGPPNAIKSCPPPSAAAAAFADKWDGTGPAAEFFSSAKKEPAFYHOGYAAAGPY 240
QY 149 HTLAPAPTY-----GVGAMNAFAPLTDKTRSHADVDGLVLSIQASLYQGYN 198
DB 241 HHN-QPVPGYLDMPVVPGLGGPG-----ESRHEPLGLPMESYQPMALPNGWN 286

RESULT 14
CH36_CERCA

```

ID CH36_CERCA STANDARD: PRT; 320 AA.
AC P17110;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Chorion protein S36.
GN CP36 OR S36.
OS Ceratitis capitata (Mediterranean fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Tephriticoidea; Tephritidae; Ceratitis.
OC NCBI_TaxID=7213;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90245561; PubMed=1692403;
RA Konojaki M., Konitopoulou K., Tollas P.P., King D.L., Swimmer C.,
RA Kataros F.C.;
RT "The chorion genes of the medfly, Ceratitis capitata, I: Structural
RT and regulatory conservation of the 836 gene relative to two
RT Drosophila species."
RL Nucleic Acids Res. 18:1731-1737(1990).
CC -I- DOMAIN: THE TETRAPEPTIDE (A-A-P-[AV]) REPEATS FOUND THROUGHOUT THE
CC PROTEIN ARE ALSO PRESENT IN MANY PROTEINS CONSTITUTING THE
CC PROTECTIVE ENVELOPE OF OTHER SPECIES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X51342; CAA35723.1; -.
DR PIR: S09208; S09208.
DR Chorion; Repeat.
FT REPEAT 178 181 1.
FT REPEAT 258 261 2.
FT REPEAT 266 269 3.
FT REPEAT 274 277 4.
FT REPEAT 280 290 5.
FT REPEAT 320 AA; 32319 MW; ECF9B72FFBE54C61 CRC64;
SQ SEQUENCE 320 AA; 32319 MW; ECF9B72FFBE54C61 CRC64;

Query Match 8.6%; Score 93.5; DB 1; Length 320;
Best Local Similarity 24.7%; Pred. No. 1.2;
Matches 40; Conservative 11; Mismatches 48; Indels 63; Gaps 6;

OY 4 KKAANPYTNGMKLNPVGAIVSP-----DFYAGTVLLQANDGSSMTSGPSSLVYTSA 57
DB 193 QGVINKVPTPLSLNPVYVYKPKKIDAPLVGVQONYPGSSYSAP----- 244
OY 58 MGFPPYPAATTAAYAGAILRGGRVYNTFRPAAPRPPIPAAGVYVQEPVYGNKLLQG 117
DB 245 -----AASVE-----PAPPSYSAAPAO----- 262
OY 118 GYAAYRYAOPTATAAAYSDSYGVYADPYHHTLAPAPTYG 159
DB 263 ----SYN-AAPAPSYSAAPASYG----AAPSASYDAPAPASYG 297

RESULT 15
GAT6_CHICK STANDARD: PRT; 387 AA.
AC P43633;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription factor GATA-6 (GATA binding factor-6).
GN GATA6.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.

```

```

OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=white leghorn;
RX MEDLINE=94365018; PubMed=8083222;
RA Laverriere A.C., Macneill C., Mueller C., Poelmann R.E.,
RA Butch J.B.E., Evans T.;
RT "GATA-4/5/6, a subfamily of three transcription factors transcribed
RT in developing heart and gut."
RT J. Biol. Chem. 269:23177-23184(1994).
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC -I- TISSUE SPECIFICITY: MORE ABUNDANT IN STOMACH, AND IN SMALL
CC INTESTINE. LOWER LEVELS IN LUNG, LIVER, OVARY AND HEART.
CC -----
CC -I- SIMILARITY: CONTAINS 2 GATA-TYPE ZINC FINGERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U11889; AAA57505.1; -.
DR HSPB; P17679; IGF.
DR TRANSFAC; T02684; -.
DR InterPro; IPR000679; Znf_GATA.
DR InterPro; IPR001164; hrip_like.
DR Pfam; PF00320; GATA; 2.
DR PRINTS; PR00619; GATAZNFINGER.
DR SMART; SM00401; ZNF_GATA; 2.
DR PROSITE; PS00344; GATA_ZN_FINGER_1; 2.
DR PROSITE; PS50114; GATA_ZN_FINGER_2; 2.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW Nuclear protein.
FT ZN_FING 181 205 GATA-TYPE 1.
FT ZN_FING 235 259 GATA-TYPE 2.
FT DOMAIN 240 POLY-THR.
SQ SEQUENCE 387 AA; 40249 MW; 6EEF457BB34DB2C0 CRC64;

Query Match 8.6%; Score 93; DB 1; Length 387;
Best Local Similarity 26.0%; Pred. No. 1.6;
Matches 47; Conservative 15; Mismatches 51; Indels 68; Gaps 10;

OY 20 VCAVYPPFYACTVLLQANDGSSMYSGPSSLVYT-----SAMPGFY----- 63
DB 7 IASQGPARYDGS-----FGFTMSAPSSPYVPTTRVGSVLPILPYLQGGAAQPG 58
OY 64 -----PAATAAA--AYRGAILRGGRVYNTFRPAAPRPPI-----PAYGVVYQ 106
DB 59 HAPAGHWSQPAAESPSYGAAGAPSGR-----FPYSASPPVANGASREQYGGGLAA 111
OY 107 EPVYG--NKLQGGYAARYAOPRTATAAY--SDSYGRVYADPYHHTL-----APA 155
DB 112 RQYQGLPRLPLNGSY-----PAPYASYVQPOLGPMWPAAPFENSVLHCLQGRAAI 162
OY 156 P 156
DB 163 P 163

Search completed: March 12, 2003, 22:24:41
Job time : 19 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Cupugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2003, 22:22:11 ; Search time 43 Seconds
(without alignments)
972.734 Million cell updates/sec

Title: US-09-809-545A-2

Perfect score: 1086
Sequence: 1 MTKKAVNPYTNCKMLNPV.....VLSSLQASIVQGGYRFPAY 203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: SPTRMBL_21:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_prodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*
16: sp_virus:*
17: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1063	97.9	418 4 Q8WYB2	Q8WYB2 homo sapien
2	1061	97.7	417 11 Q8R4Z7	Q8R4Z7 mus musculu
3	912	84.0	392 4 Q8TAE3	Q8TAE3 homo sapien
4	729.5	67.2	395 4 Q8TAE2	Q8TAE2 homo sapien
5	676	62.2	376 6 Q95KI0	Q95KI0 macaca fasc
6	532.5	49.0	293 4 Q9UGW3	Q9UGW3 homo sapien
7	526	48.4	380 4 Q8WYB1	Q8WYB1 homo sapien
8	526	48.4	390 4 Q8TUD0	Q8TUD0 homo sapien
9	525	48.3	377 11 Q923W8	Q923W8 mus musculu
10	525	48.3	377 11 Q8V162	Q8V162 mus musculu
11	522	48.1	358 4 Q96DZ6	Q96DZ6 homo sapien
12	519	47.8	450 4 Q96N17	Q96N17 mus musculu
13	456	42.0	303 11 Q8V161	Q8V161 mus musculu
14	351	32.3	263 4 Q8TCM0	Q8TCM0 homo sapien
15	296	27.3	435 11 Q8R2T5	Q8R2T5 mus musculu
16	129	11.9	367 5 Q9VT99	Q9VT99 drosophila

17	123.5	11.4	326 5 Q9NFX5	Q9NFX5 ceratitis c
18	116.5	10.7	326 5 Q62010	Q62010 ceratitis c
19	110.5	10.2	219 5 Q9V875	Q9V875 drosophila
20	108.5	10.0	291 4 Q92567	Q92567 homo sapien
21	104.5	9.6	226 5 Q9SV16	Q9SV16 myzus persi
22	103	9.5	332 5 Q9NFX7	Q9NFX7 ceratitis c
23	103	9.5	338 16 Q9R233	Q9R233 delnoccocus
24	102	9.4	604 16 Q8XPW5	Q8XPW5 ralsotonia s
25	101.5	9.3	950 5 Q9VE45	Q9VE45 drosophila
26	100	9.2	830 5 Q9W176	Q9W176 drosophila
27	99	9.1	542 5 Q23446	Q23446 caenorhabdi
28	98.5	9.1	723 11 Q9YK84	Q9YK84 mus musculu
29	98	9.0	316 5 Q16463	Q16463 caenorhabdi
30	97.5	9.0	332 5 Q62009	Q62009 ceratitis c
31	97.5	9.0	358 10 Q94223	Q94223 oryza sativ
32	97.5	9.0	828 5 Q9M349	Q9M349 drosophila
33	97	8.9	179 5 Q9VV18	Q9VV18 drosophila
34	97	8.9	363 5 Q01353	Q01353 drosophila
35	97	8.9	384 16 Q9FCU3	Q9FCU3 streptomyce
36	97	8.9	723 11 Q9WVH2	Q9WVH2 mus musculu
37	97	8.9	811 16 Q9X8T0	Q9X8T0 streptomyce
38	97	8.9	835 5 Q95YR9	Q95YR9 leishmania
39	96.5	8.9	131 10 Q40415	Q40415 nicotiana s
40	96.5	8.9	748 10 Q9T0G5	Q9T0G5 arabidopsis
41	96.5	8.9	1024 12 Q89815	Q89815 bovine herp
42	96	8.8	200 5 Q961V4	Q961V4 drosophila
43	95.5	8.8	205 5 P82167	P82167 locusta mig
44	95.5	8.8	392 5 Q9BLS3	Q9BLS3 leishmania
45	95	8.7	275 5 Q9BP11	Q9BP11 caenorhabdi

ALIGNMENTS

RESULT 1
ID Q8WYB2 PRELIMINARY; PRT; 418 AA.
AC Q8WYB2;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hexaribonucleotide binding protein 1 isoform gamma.
GN HRNBPL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen W., Winkelman J.C.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Chen W., Chu Z.-L., Blough R.I., Liu L., Hoppes B., Winkelman J.C.;
RT "Molecular Cloning and Chromosomal Localization of a Novel Human Brain, Heart and Skeletal Muscle Specific RNA Binding Protein Gene Homologous to fox-1 in Caenorhabditis elegans.";
RT Homologous to fox-1 in Caenorhabditis elegans.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Chen W., Winkelman J.C.;
RL "The Exon-Intron Organization of the Human HRNBPL Gene.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF229057; AAL71904.1; -;
DR EMBL; AF109120; AAL83409.1; -;
DR EMBL; AF109107; AAL83409.1; JOINED.
DR EMBL; AF109108; AAL83409.1; JOINED.
DR EMBL; AF109109; AAL83409.1; JOINED.
DR EMBL; AF109110; AAL83409.1; JOINED.
DR EMBL; AF109111; AAL83409.1; JOINED.
DR EMBL; AF109112; AAL83409.1; JOINED.
DR EMBL; AF109113; AAL83409.1; JOINED.
DR EMBL; AF109114; AAL83409.1; JOINED.


```

AC 08TAF2;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hexayribonucleotide binding protein 1 isoform beta.
GN HNRNP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen W., Winkelman J.C.;
RT "The Exon-Intron Organization of the Human HNRNP1 Gene.";
RT Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF109106; AAL83406.1; -.
DR EMBL; AF109120; AAL83407.1; JOINED.
DR EMBL; AF109107; AAL83407.1; JOINED.
DR EMBL; AF109108; AAL83407.1; JOINED.
DR EMBL; AF109109; AAL83407.1; JOINED.
DR EMBL; AF109110; AAL83407.1; JOINED.
DR EMBL; AF109111; AAL83407.1; JOINED.
DR EMBL; AF109112; AAL83407.1; JOINED.
DR EMBL; AF109113; AAL83407.1; JOINED.
DR EMBL; AF109114; AAL83407.1; JOINED.
DR EMBL; AF109115; AAL83407.1; JOINED.
DR EMBL; AF109116; AAL83407.1; JOINED.
DR EMBL; AF109117; AAL83407.1; JOINED.
DR EMBL; AF109118; AAL83407.1; JOINED.
DR EMBL; AF109119; AAL83407.1; JOINED.
SQ SEQUENCE 395 AA; 42402 MW; 2299F9127A1DE39 CRC64;

Query Match 67.2%; Score 729.5; DB 4; Length 395;
Best Local Similarity 81.2%; Pred. No. 3.2e-52;
Matches 143; Conservative 2; Mismatches 12; Indels 19; Gaps 2;

OY 1 MTNKKAVNPYNGMKLNPNVGAIVSPDPFAGTVLLCQANQSGSSMYSGPSSLVYTSAMPG 60
DB 216 MTNKKAVNPYNGMKLNPNVGAIVSPDPFAGTVLLCQANQSGSSMYSGPSSLVYTSAMPG 275

OY 61 PPYPATAAAYRGALHGRGRTVYNTFRRAAPPPIPAYGVVYQEPYGNKLLQGGYA 120
DB 276 PPYPATAAAYRGALHGRGRTVYNTFRRAAPPPIPAYGVVYQEPYGNKLLQGGYA 335

OY 121 AYRYAQPPTPATAAAYSDSYGRVY-AAAPYH-----HTLAPAPT 157
DB 336 AYRYAQPPTPATAAAYSDSYGRVY-AAAPYH-----HTLAPAPT 391

RESULT 5
OY 095K10 PRELIMINARY; PRT; 376 AA.
AC 095K10;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical 40.6 kDa protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RA Tissus-TEMPORAL LOBE RIGHT;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,

```

```

RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries.";
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB060859; BAB46877.1; -.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; Rrm; 1.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 376 AA; 40565 MW; DAA7A4DD07B1030A CRC64;

Query Match 62.2%; Score 676; DB 6; Length 376;
Best Local Similarity 76.7%; Pred. No. 7.6e-48;
Matches 135; Conservative 4; Mismatches 17; Indels 20; Gaps 3;

OY 1 MTNKKAVNPYNGMKLNPNVGAIVSPDPFAGTVLLCQANQSGSSMYSGPSSLVYTSAMPG 60
DB 198 MTNKKAVNPYNGMKLNPNVGAIVSPDPFAGTVLLCQANQSGSSMYSGPSSLVYTSAMPG 257

OY 61 PPYPATAAAYRGALHGRGRTVYNTFRRAAPPPIPAYGVVYQEPYGNKLLQGGYA 120
DB 258 PPYPATAAAYRGALHGRGRTVYNTFRRAAPPPIPAYGVVYQEPYGNKLLQGGYA 316

OY 121 AYRYAQPPTPATAAAYSDSYGRVY-AAAPYH-----HTLAPAPT 157
DB 317 AYRYAQPPTPATAAAYSDSYGRVY-AAAPYH-----HTLAPAPT 372

RESULT 6
OY 09UGW3 PRELIMINARY; PRT; 293 AA.
AC 09UGW3;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE D41P2.2 (Supported by GENSCAN) (Fragment).
GN RBM9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Whiteley M.;
RT Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL049748; CAB63055.1; -.
DR HSSP; P11940; IGVJ.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; Rrm; 1.
DR SMART; SMO0360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
FT NON TER 1
SQ SEQUENCE 293 AA; 30983 MW; F9A823DD70CC1004 CRC64;

Query Match 49.0%; Score 532.5; DB 4; Length 293;
Best Local Similarity 52.8%; Pred. No. 3.6e-36;
Matches 114; Conservative 17; Mismatches 44; Indels 41; Gaps 7;

OY 1 MTNKKAVNPYNGMKLNPNVGAIVSPDPFAGTVLLCQANQSGSSMYSGPSSLVYTSAMPG 56
DB 106 MTNKKAVNPYNGMKLNPNVGAIVSPDPFAGTVLLCQANQSGSSMYSGPSSLVYTSAMPG 165

OY 57 AMGPFPY-AAATAAAYGAHLRGRGRTVYNTFRRAAPPPIPAYGVVYQEPYGNKLL 115
DB 166 IIPGPYPTAATTAATAAFAFGAHLRGRGRTVYNTFRRAAPPPIPAYGVVYQEPYGNKLL 224

OY 116 OGGVAAVRYAQPPTPAT-----AAAYSDSYGRVYADPPYHHTLAPAPTYGVGMNFA 167
DB 225 OGGVAAVRYAQPPTPATATAAATAAAYSDSYGRVYADPPYHHTLAPAPTYGVGMNFA 279

OY 168 PLTDAKTRSHADVDGLVLSLQASLYOGVNRFAFY 203

```

Db 280 -----ASLYRGYSRFPAY 293

RESULT 7

Q8WYB1 PRELIMINARY; PRT; 380 AA.

AC Q8WYB1
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hexarionnucleotide binding protein 2 (RNA binding motif protein 9).
 GN HRNP2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euteria; Primates; Catarrhini; Homidae; Homo.
 NCBI_TaxId:9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Chen W., Winkelman J.C.;
 RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RC TISSUE-UTERUS;
 RA Strausberg R.;
 RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; BC025281; AAL71905.1; -
 DR InterPro; IPR00104; Antifreeze_1.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; rrm; 1.
 DR PRINTS; PRO0308; ANTIFREEZE.
 DR SMART; SM00360; RRM; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
 DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
 SQ SEQUENCE 380 AA; 40351 MW; 601P4E31E7C20715 CRC64;

Query Match 48.4%; Score 526; DB 4; Length 380;
 Best Local Similarity 52.7%; Pred. No. 1.7e-35;
 Matches 116; Conservative 15; Mismatches 43; Indels 46; Gaps 8;

QY 1 MTNKA VNPYTGWKLNPVGA VYSPDYAGTVLLCOAN--OEGSSMVGSPSS-----L 52
 DB 190 MTNKKVNTVYANGWKLSPVGA VYGPDELTAASSFQADVSLGNDAAVPLSGRGGINTYIP 249
 QY 53 VYTSAMPGEPPY-AAATAAAAYRGALRGGRVTYNTFRAAAPPPPIPAVGVVYQEPVYG 111
 DB 250 ISLPLVPGFPYPTAATTAATAAFRGALHGRGRVYGA VR-AVPEPTAIPAVGVVYQDGFYG 308
 QY 112 NKLLOGVAAYRYAQPPTPAT-----AAAYS DSYGRVYAADPYHHTLAPAPTYGVGAM 163
 DB 309 AD-LVGGYAAYRYAQPATTAATAAATAAAAYSDYGRVYTDAPY-HALAPAA SYGVGAV 366
 QY 164 NAFAPLTDKTRSHADVGVLVSLQASLYOGGYNRFAPY 203
 DB 367 -----ASLYRGYSRFPAY 380
 RESULT 8
 Q8TD00 PRELIMINARY; PRT; 390 AA.
 AC Q8TD00;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE RNA binding motif protein 9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euteria; Primates; Catarrhini; Homidae; Homo.
 NCBI_TaxId:9606;
 RN (1)
 RP SEQUENCE FROM N.A.

RX MEDLINE=21864496; PubMed=11875103;
 RA Norris J.D., Fan D., Sheik A., McDonnell D.P.;
 RT "A Negative Coregulator for the Human ER.";
 RL Mol. Endocrinol. 16:459-468(2002).
 DR EMBL; AY072786; AAL67150.1; -
 SQ SEQUENCE 390 AA; 41374 MW; 40D6C8C40764B317 CRC64;

Query Match 48.4%; Score 525; DB 11; Length 377;
 Best Local Similarity 53.7%; Pred. No. 2e-35;
 Matches 116; Conservative 14; Mismatches 44; Indels 42; Gaps 8;

QY 1 MTNKA VNPYTGWKLNPVGA VYSPDYAGTVLLCOAN--OEGSSMVGSPSS-----L 52
 DB 200 MTNKKVNTVYANGWKLSPVGA VYGPDELTAASSFQADVSLGNDAAVPLSGRGGINTYIP 259
 QY 53 VYTSAMPGEPPY-AAATAAAAYRGALRGGRVTYNTFRAAAPPPPIPAVGVVYQEPVYG 111
 DB 260 ISLPLVPGFPYPTAATTAATAAFRGALHGRGRVYGA VR-AVPEPTAIPAVGVVYQDGFYG 318
 QY 112 NKLLOGVAAYRYAQPPTPAT-----AAAYS DSYGRVYAADPYHHTLAPAPTYGVGAM 163
 DB 319 AD-LVGGYAAYRYAQPATTAATAAATAAAAYSDYGRVYTDAPY-HALAPAA SYGVGAV 376
 QY 164 NAFAPLTDKTRSHADVGVLVSLQASLYOGGYNRFAPY 203
 DB 377 -----ASLYRGYSRFPAY 390
 RESULT 9
 Q923W8 PRELIMINARY; PRT; 377 AA.
 AC Q923W8;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative RNA-binding protein fwh.
 GN FWH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxId:10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA MEDLINE=21294761; PubMed=11401487;
 RA Lieberman A.P., Friedlich D.L., Harrison G., Howell B.W., Jordan C.L., Breedlove S.M., Fischbeck K.H.;
 RT "Androgens Regulate the Mammalian Homologues of Invertebrate Sex Determination Genes tra-2 and fox-1.";
 RL Biochem. Biophys. Res. Commun. 282:499-506(2001).
 DR EMBL; AF387322; AAK64287.1; -
 DR MGD; MGI:1933973; Fwh.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; rrm; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
 DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
 SQ SEQUENCE 377 AA; 40165 MW; 249FA8A6A0D155A CRC64;

Query Match 48.3%; Score 525; DB 11; Length 377;
 Best Local Similarity 53.7%; Pred. No. 2e-35;
 Matches 116; Conservative 14; Mismatches 44; Indels 42; Gaps 8;

QY 1 MTNKA VNPYTGWKLNPVGA VYSPDYAGTVLLCOAN--OEGSSMVGSPSS--VYTS 56
 DB 191 MTNKKVNTVYANGWKLSPVGA VYGPDELTAASSFQADVSLGNDAAVPLSGRGGINTYIP 250
 QY 57 AMPGFPY-AAATAAAAYRGALRGGRVTYNTFRAAAPPPPIPAVGVVYQEPVYGKLL 115
 DB 251 ILPGFPYPTAATTAATAAFRGALHGRGRVYGA VR-AVPEPTAIPAVGVVYQDGFYAD-L 308
 QY 116 OGSYAAYRYAQPPTPAT-----AAAYS DSYGRVYAADPYHHTLAPAPTYGVGAMNAPA 167
 DB 309 YGSYAAYRYAQPATTAATAAATAAAAYSDYGRVYTDAPY-HALAPAA SYGVGAV---- 363

QY 168 PLTDKTRSHADVGLVLSLQASTIYOGGYNRFAPY 203
DB 364 -----ASLYRGYSRFPAY 377

RESULT 10
QY 08VI62 PRELIMINARY; PRT; 377 AA.
AC 08VI62;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hexabinucleotide binding protein 2.
GN FXH OR HRNBP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Chen W., Winkelman J.C.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF229055; AAL71902.1; -.
DR MGD; MG1:1933973; Fxh.
DR InterPro; IPR000104; Antifreeze_1.
DR Pfam; PF00076; rtm; 1.
DR PRINTS; PRO0308; ANTIFREEZE1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PSS0102; RRM; 1.
DR PROSITE; PS00030; RRM RNP 1; UNKNOWN 1.
SQ SEQUENCE 377 AA; 40183 MW; 249E72CAFA7C505A CRC64;

Query Match 48.3%; Score 525; DB 11; Length 377;
Best Local Similarity 53.7%; Pred. No. 2e-35;
Matches 115; Conservative 14; Mismatches 44; Indels 42; Gaps 8;

QY 1 MTNKAVPYNGMKLNPVGVAVSPDFYAGTVLLCOAN--QEGSSMWGSPSS--VYTS 56
DB 191 MTNKMVTTPYANGKLSPVGVAVGPELYAASSFQADVSLGNDAVPLSGGINTYIPL 250
QY 57 AMGFPYP-AATAAAYRGALHGRGRTVNTFRAADPPPIPAYGVVYQEPYGNKL 115
DB 251 IIPGFYPPTAATTAAFRGALHGRGRTVYAVR-AVPTAIIPAYPGVYQDGFYGD-L 308
QY 116 QGGYAAARYAOPPTAT-----AAASDSYGRVYAADPYHTLAPRYGVGAMNAFA 167
DB 309 YGGYAAARYAOPATATATAAATAAASDYGRTYADPY-HALAPASVGVAV---- 363
QY 168 PLTDKTRSHADVGLVLSLQASTIYOGGYNRFAPY 203
DB 364 -----ASLYRGYSRFPAY 377

RESULT 11
QY 096DZ6 PRELIMINARY; PRT; 358 AA.
AC 096DZ6;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Similar to fox-1 homolog (C. elegans).
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tissue=ADRENAL GLAND;
RA Straussberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013115; AAH13115.1; -.
DR InterPro; IPR000504; RNA_rec_mot.

DR Pfam; PF00076; rtm; 1.
DR PROSITE; PSS0102; RRM; 1.
SQ SEQUENCE 358 AA; 37907 MW; BAE3740C178E441 CRC64;

Query Match 48.1%; Score 522; DB 4; Length 358;
Best Local Similarity 53.2%; Pred. No. 3.4e-35;
Matches 115; Conservative 15; Mismatches 44; Indels 42; Gaps 8;

QY 1 MTNKAVPYNGMKLNPVGVAVSPDFYAGTVLLCOAN--QEGSSMWGSPSS--VYTS 56
DB 172 MTNKMVTTPYANGKLSPVGVAVGPELYAASSFQADVSLGNDAVPLSGGINTYIPL 231
QY 57 AMGFPYP-AATAAAYRGALHGRGRTVNTFRAADPPPIPAYGVVYQEPYGNKL 115
DB 232 IIPGFYPPTAATTAAFRGALHGRGRTVYAVR-AVPTAIIPAYPGVYQDGFYGD-L 289
QY 116 QGGYAAARYAOPPTAT-----AAASDSYGRVYAADPYHTLAPRYGVGAMNAFA 167
DB 290 YGGYAAARYAOPATATATAAATAAASDYGRTYADPY-HALAPASVGVAV---- 344
QY 168 PLTDKTRSHADVGLVLSLQASTIYOGGYNRFAPY 203
DB 345 -----ASLYRGYSRFPAY 358

RESULT 12
QY 096NL7 PRELIMINARY; PRT; 450 AA.
AC 096NL7;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE CDNA FLJ30651 fts, clone D3OST2000184, highly similar to RNA binding motif protein 9.
GN Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Nimomiya K., Nagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Aoe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Kanemori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahara K., Maehno Y., Negai K., Isogai T.,
RL "NEDO human cDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK055213; BAB70875.1; -.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rtm; 1.
DR PROSITE; PSS0102; RRM; 1.
DR PROSITE; PS00030; RRM RNP 1; UNKNOWN 1.
SQ SEQUENCE 450 AA; 47236 MW; 173D0C55D5CD762 CRC64;

Query Match 47.8%; Score 519; DB 4; Length 450;
Best Local Similarity 52.3%; Pred. No. 7.8e-35;
Matches 115; Conservative 14; Mismatches 45; Indels 46; Gaps 8;

QY 1 MTNKAVPYNGMKLNPVGVAVSPDFYAGTVLLCOAN--QEGSSMWGSPSS-----L 52
DB 260 MTNKMVTTPYANGKLSPVGVAVGPELYAASSFQADVSLGNDAVPLSGGINTYIPL 319
QY 53 VYTSAMGFPYP-AATAAAYRGALHGRGRTVNTFRAADPPPIPAYGVVYQEPYGN 111
DB 320 ISLPLVPGFPYPPTAATTAAFRGALHGRGRTVYAVR-AVPTAIIPAYPGVYQDGFYGD 378
QY 112 NKLOGGYAAARYAOPPTAT-----AAASDSYGRVYAADPYHTLAPRYGVGAM 163
DB 379 AD-LYGGYAAARYAOPATATATAAATAAASDYGRTYADPY-HALAPASVGVAV 436

QY 164 NAFAPLDAKTRSHADDVGLVSSLOASTYOCGYNRFAPY 203
 Db 437 -----ASLYRGYSRFAFY 450

RESULT 13

Q8VI61 PRELIMINARY; PRT; 303 AA.
 AC Q8VI61;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hexaribonucleotide binding protein 3 (Fragment).
 GN HRNBP3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BAJB/C;
 RA Chen W., Winkelman J.C.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF229056; AAL71903.1; -.
 DR InterPro; IPR005054; RNA_rec_mot.
 DR Pfam; PF00076; rrm; 1.
 DR SMART; SM00360; RRM; 1.
 DR PROSITE; PS0102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
 FT NON_TER 1
 SQ SEQUENCE 303 AA; 32864 MW; 74B0AF9F9FB7E935 CRC64;

Query Match 42.0%; Score 456; DB 11; Length 303;
 Best Local Similarity 45.5%; Pred. No. 7.5e-30;
 Matches 95; Conservative 6; Mismatches 32; Indels 76; Gaps 4;

QY 3 NKAANVPYTNMGKLNIPVGAIVSPDFYAGTVLLCOANOGSSMYSGPSSLYVTSAMPGP 62
 Db 123 NNATARVMTNGMKLNIPVGVGVGPEFYAVT-----SFP 155
 QY 63 YAAATAAAYRGAHLRGRRTVYNTFRAPAPPPPIPAYGG----- 102
 Db 156 YTTGTAAVAVRGTHLGRGAAYNTFRAPPPPIPTTGAALEQTLVKNPVPAGIAPCP 215
 QY 103 -----VYQEPVYGNKLLGGYAAVRYAOPTPATAAV 135
 Db 216 LPPQTPPEAPYTPSPAFPPPLSCFPASRVVYQDGFYGAE-IYGGYAAVRYAQAAPATAAV 274
 QY 136 SDSYGRVY-AADPYHHTLAPAPTYGVGM 163
 Db 275 SDSYGRVYAADPYHHTIGPTATYSIGTM 303

RESULT 14

Q8TCM0 PRELIMINARY; PRT; 263 AA.
 AC Q8TCM0;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 28.6 kDa protein (Fragment).
 GN DKFZP547L059.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRN;
 RA Blocker H., Boecher M., Brande P., Mewes H.W., Weij B., Wiemann S.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL713700; CAD28499.1; -.
 KW Hypothetical protein.

FT NON_TER 263 263
 SQ SEQUENCE 263 AA; 28565 MW; 56C2D8538468CE1 CRC64;

Query Match 32.3%; Score 351; DB 4; Length 263;
 Best Local Similarity 95.6%; Pred. No. 2.8e-21;
 Matches 65; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTNKAANVPYTNMGKLNIPVGAIVSPDFYAGTVLLCOANOGSSMYSGPSSLYVTSAMPGP 60
 Db 196 MTNKTAVPYTNMGKLNIPVGAIVSPDFYAGTVLLCOANOGSSMYSA PSSLYVTSAMPGP 255
 QY 61 FFPYATA 68
 Db 256 FFPYATA 263

RESULT 15

Q8R2T5 PRELIMINARY; PRT; 435 AA.
 AC Q8R2T5;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Similar to RNA binding motif protein 9.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Straubeberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC027263; AAH27263.1; -.
 SQ SEQUENCE 435 AA; 46252 MW; 62A355605FF8A5B4 CRC64;

Query Match 27.3%; Score 296; DB 11; Length 435;
 Best Local Similarity 50.4%; Pred. No. 1.7e-16;
 Matches 70; Conservative 10; Mismatches 53; Indels 6; Gaps 4;

QY 1 MTNKAANVPYTNMGKLNIPVGAIVSPDFYAGTVLLCOAN--QEGSSMYSGPSSL--VYTS 56
 Db 259 MTNKGAVTPYANGMKLSPVGAIVSPDFYAGTVLLCOANVPLSGRGIGINTYIPL 318
 QY 57 AMPGPPYP-AATAAAYRGAHLRGRRTVYNTFRAPAPPPPIPAYGVYQEPVYGNKLL 115
 Db 319 IIPGPPYPYATTAATAAFRAHLRGRRTVYGAVR-AVPEYALPAPGVDMQPTDMSLL 377
 QY 116 OGGYAAVRYAOPTPATAA 134
 Db 378 QPQPOLQPLQPLTATVTA 396

Search completed: March 13, 2003, 00:54:09
 Job time : 44 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2003, 22:24:21 ; Search time 19 Seconds
(without alignments)

314.361 Million cells updates/sec

Title: US-09-809-545A-2

Perfect score: 1086

Sequence: 1 MTNKKAVNPYTNQWKLNPV.....VLSSLSQASIVQGGYRFPAPY 203

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents AA.*

2: /cgn2_6/ptodata/1/iaa/5A COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/5B COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/6A COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/6B COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/6C COMB.pep.*

7: /cgn2_6/ptodata/1/iaa/6D COMB.pep.*

8: /cgn2_6/ptodata/1/iaa/6E COMB.pep.*

9: /cgn2_6/ptodata/1/iaa/6F COMB.pep.*

10: /cgn2_6/ptodata/1/iaa/6G COMB.pep.*

11: /cgn2_6/ptodata/1/iaa/6H COMB.pep.*

12: /cgn2_6/ptodata/1/iaa/6I COMB.pep.*

13: /cgn2_6/ptodata/1/iaa/6J COMB.pep.*

14: /cgn2_6/ptodata/1/iaa/6K COMB.pep.*

15: /cgn2_6/ptodata/1/iaa/6L COMB.pep.*

16: /cgn2_6/ptodata/1/iaa/6M COMB.pep.*

17: /cgn2_6/ptodata/1/iaa/6N COMB.pep.*

18: /cgn2_6/ptodata/1/iaa/6O COMB.pep.*

19: /cgn2_6/ptodata/1/iaa/6P COMB.pep.*

20: /cgn2_6/ptodata/1/iaa/6Q COMB.pep.*

21: /cgn2_6/ptodata/1/iaa/6R COMB.pep.*

22: /cgn2_6/ptodata/1/iaa/6S COMB.pep.*

23: /cgn2_6/ptodata/1/iaa/6T COMB.pep.*

24: /cgn2_6/ptodata/1/iaa/6U COMB.pep.*

25: /cgn2_6/ptodata/1/iaa/6V COMB.pep.*

26: /cgn2_6/ptodata/1/iaa/6W COMB.pep.*

27: /cgn2_6/ptodata/1/iaa/6X COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	551	50.7	330	US-09-145-391-2	Sequence 2, Appli
2	288	26.5	366	US-09-176-657-2	Sequence 2, Appli
3	96	8.8	760	US-08-195-152-2	Sequence 2, Appli
4	94	8.7	349	US-09-343-011B-1	Sequence 1, Appli
5	93	8.6	444	US-09-252-292C-27	Sequence 27, Appli
6	92.5	8.5	362	US-08-437-027-21	Sequence 21, Appli
7	92.5	8.5	365	US-08-437-027-20	Sequence 20, Appli
8	92.5	8.5	656	US-08-343-443B-2	Sequence 2, Appli
9	92.5	8.5	656	US-09-214-564A-4	Sequence 4, Appli
10	91	8.4	346	US-09-343-011B-2	Sequence 2, Appli
11	87.5	8.1	980	US-09-442-100-8	Sequence 8, Appli
12	87	8.0	379	US-08-552-142A-11	Sequence 11, Appli
13	86.5	8.0	262	US-08-946-914-14	Sequence 14, Appli
14	86.5	8.0	262	US-09-656-450-14	Sequence 14, Appli
15	86.5	8.0	4302	US-08-658-136-5	Sequence 5, Appli
16	86.5	8.0	4302	US-09-052-469-8	Sequence 8, Appli
17	86.5	8.0	4339	US-09-052-469-6	Sequence 6, Appli
18	86	7.9	207	US-09-336-536-50	Sequence 50, Appli
19	86	7.9	245	US-09-336-536-48	Sequence 48, Appli
20	85	7.8	275	US-08-431-387-3	Sequence 3, Appli
21	85	7.8	275	US-08-322-677A-7	Sequence 7, Appli
22	85	7.8	275	US-08-322-676-7	Sequence 7, Appli
23	85	7.8	275	US-08-460-343B-74	Sequence 74, Appli
24	85	7.8	275	US-08-398-028B-74	Sequence 74, Appli
25	85	7.8	275	US-08-504-265B-90	Sequence 90, Appli
26	85	7.8	275	US-08-140-083A-9	Sequence 9, Appli
27	85	7.8	275	US-08-865-203-8	Sequence 8, Appli

28	85	7.8	275	2	US-09-135-658-3	Sequence 3, Appli
29	85	7.8	275	2	US-07-849-420-8	Sequence 8, Appli
30	85	7.8	275	3	US-08-898-218-7	Sequence 7, Appli
31	85	7.8	275	3	US-08-848-793-7	Sequence 7, Appli
32	85	7.8	275	4	US-09-253-854-8	Sequence 8, Appli
33	85	7.8	275	4	US-08-955-424-8	Sequence 8, Appli
34	85	7.8	275	4	US-09-178-155-3	Sequence 3, Appli
35	85	7.8	275	4	US-09-445-270-2	Sequence 2, Appli
36	85	7.8	275	4	US-09-467-536A-2	Sequence 2, Appli
37	85	7.8	275	4	US-09-234-957-2	Sequence 2, Appli
38	85	7.8	275	4	US-08-394-011-1	Sequence 1, Appli
39	85	7.8	275	4	US-08-397-329-1	Sequence 1, Appli
40	85	7.8	275	5	PCT-US93-00418-3	Sequence 3, Appli
41	85	7.8	275	5	PCT-US95-04760-2	Sequence 2, Appli
42	85	7.8	352	1	US-07-923-260A-2	Sequence 2, Appli
43	85	7.8	352	1	US-07-923-260A-2	Sequence 3, Appli
44	85	7.8	380	1	US-08-552-142A-17	Sequence 17, Appli
45	85	7.8	382	4	US-09-255-502-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-09-145-391-2
Sequence 2, Application US/09145391
Patent No. 6194171
GENERAL INFORMATION:
APPLICANT: Pulist, Stefan M.
APPLICANT: Shibata, Hiroki
TITLE OF INVENTION: Nucleic Acids Encoding Abaxin-2 Binding Proteins,
FILE REFERENCE: CE 3093
CURRENT APPLICATION NUMBER: US/09/145,391
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 330
TYPE: PRT
ORGANISM: Homo sapiens
US-09-145-391-2

Query Match      50.7%; Score 551; DB 4; Length 330;
Best Local Similarity 97.1%; Pred. No. 2,6e-48;
Matches 102; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 MTNKKAVNPYTNQWKLNPVGVAVSPDFYAGTVLLCQANQEGSSMYSGPSSLVYTSAMP 60
Db      196 MTNKKAVNPYTNQWKLNPVGVAVSPDFYAGTVLLCQANQEGSSMYSGPSSLVYTSAMP 255

Qy      61 PPYPATTAAYRGAAHLGGRGTYYNTFRRAAPPPIRAGGVY 105
Db      256 PPYPATTAAYRGAAHLGGRGTYYNTFRRAAPPPIRAGGVY 300

RESULT 2
US-09-176-657-2
Sequence 2, Application US/09176657
Patent No. 6020164
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Tang, Y. Tom
APPLICANT: Corley, Neil C.
APPLICANT: Guejter, Karl J.
APPLICANT: Lu, Aina
APPLICANT: Baughn, Mariah R.
TITLE OF INVENTION: HUMAN RNA BINDING PROTEINS
FILE REFERENCE: PF-0611 US
CURRENT APPLICATION NUMBER: US/09/176,657
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PERL Program
```

```

; SEQ ID NO 2
; LENGTH: 366
; TYPE: prt
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: J250374
US-09-176-657-2

```

Query Match	26.5%	Score 288;	DB 3;	Length 366;
Best Local Similarity	48.9%	Pred. No. 1.9e-21;		
Matches 68;	Conservative 11;	Mismatches 54;	Indels 6;	Gaps 4

0y 1 MTNKAANPYNMGKLNIVNVAVYSPDEFAGTVLLCQAN--QEISSVYGGSSLD--VYTS 56

Db 190 MTNKKMTPIYANGMKLSIVGAGVGPBELIYAASSFOADVSLGNDAAVPLSGRGGINITYPL 249

0y 57 AMPGFPY-AATAAANAAYGALHGRGRVTNYPFAAAPPEIPAYGVGVYQEPYVGNKLL 115

Db 250 IIPGPIYPTAALTAAAFGAHLRGRGRIVGAVR-AVPTAIPAYPGVDQPTDMHSLLL 308

0y 116 QGVAAVRYAQPATATAA 134

Db 309 QPQPLLPDLPVTVTMA 327

RESULT 3
US-08-195-152-2
; Sequence 2, Application US/08195152

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

ADDRESSEE: Flehr, Hobbach, Teet Abritton & Herberich
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: United States
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/195,152
 FILING DATE: 14-FEB-1994

```

: ATTORNEY/AGENT INFORMATION:
: NAME: Trecartin, Richard F.
: REGISTRATION NUMBER: 31,801
: REFERENCE/DOCKET NUMBER: A-59551/RT/RMS
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 760 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
:
: JS-08-195-152-2

```

```

Query Match      8.8%; Score 96; DB 1; Length 760;
Best Local Similarity 23.7%; Pred. No. 0.18;
Matches 42; Conservative 24; Mismatches 61; Indels 50; Gaps 6

QY 42 GSSMY-----SCPSSIVYTSAMPFPYPAATAAAVRGAAHLGRGRTVYNTFRRAA 92

```

```

Db      223  GSNLYGCSASNP LLDGGA VAVNNSAV-----AAAAAAYVDGKH---DYYYSNSMQYT 272
QY      93  PPPPIPAYGVVYVEPYYGK LLOGGAAATRYAOPPTATAA-----134
Db      273  PPPPSYSGTPTAAATARAQKMEPGAAAAAALITPSYASGNNNSQLYSSPYAGYNF 332
QY      135  -----YSDSYGRVVAADPDPHHHTLAPAP---TYVGV-AMNAAFAPLDAKTRSHA 178
Db      333  GQDDYGAYNEQYIGNYISPA NYISRYAVNSPSSSSASHGCHGVAAVNASNLTSSPDTDS 389

```

```

RESULT 4
US-09-343-011B-1
; Sequence 1, Application US/09343011B
; Patent No. 6300473
; GENERAL INFORMATION:
; APPLICANT: Stephane Richard
; TITLE OF INVENTION: SLM-1 AND SLM-2; NOVEL
; TITLE OF INVENTION: SAM68-LIKE MAMMALIAN PROTEINS
; FILE REFERENCE: A32561
; CURRENT APPLICATION NUMBER: US/09/343, 011B
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: CA 2265271
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-343-011B-1

```

Query Match	8.7%	Score 94;	DB 4;	Length 349;
Best Local Similarity	30.7%;	Pred. No. 0.1;		
Matches 54;	Conservative 11;	Mismatches 51;	Indels 60;	Gaps 12

QY	73	RGALHGRGGRIVYNTF	---AAAPPPPIYAGGVVVOEPIYGNKLQG	---	117
Db	186	RGGRGRGGRITPTAASRGCGAVPPPPPGRRVLPR	---GTTVHGLALPVPPIARGV	242	242
QY	118	-----GVA---YRVOPRTATA	---AYSDSYGRVYADPY	---HHTLAPPT	---
Db	243	PTPRARGAALPGRR	---APPPAHAYEYEGIDDSYGGSEYDDQTYEALDINSVYPTQSVPE	301	301
QY	158	---YGVG---ANNAFAPLIDAKTRSHADVDGLVLSLQ	---SIYCGYNNRFPY	203	203
Db	302	YYDGHGHNEDAYIVSYVAPEEWATR	---SSLKAPPPPSAGSGY	REHY	346

```

RESULT 5
US-09-252-292C-27
: Sequence 27, Application US/09252292C
: Patent No. 6245968
: GENERAL INFORMATION:
: APPLICANT: Boudec, Philippe
: APPLICANT: Rodgers, Matthew
: APPLICANT: Dumas, Florence
: TITLE OF INVENTION: Mutated hydroxyphenylpyruvate dioxygenase, DNA
: TITLE OF INVENTION: Sequence and isolation of plants which contain such a
: FILE REFERENCE: 5500*31
: CURRENT APPLICATION NUMBER: US/09/252,292C
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 08/982,772
: PRIOR FILING DATE: 1997-12-02
: NUMBER OF SEQ. ID NOS: 31
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 27
: LENGTH: 444
: TYPE: PRT
: ORGANISM: Zea mays
US-09-252-292C-27

```



```

; APPLICANT: Delattre, Olivier
; APPLICANT: Desmaze, Chantal
; APPLICANT: Meloc, Thomas
; APPLICANT: Peter, Martine
; APPLICANT: Ploegastel, Beatrice
; APPLICANT: Thomas, Gilles
; APPLICANT: Zucman, Jessica
; TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
; TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
; TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
; TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
; NUMBER OF SEQUENCES: 129
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: AEDIT 1.0 DOS text editor
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/343,443B
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00494
; FILING DATE: 19-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/06123
; FILING DATE: 20-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 989,6121P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 656 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-343-443B-2

Query Match      8.5%; Score 92.5; DB 2; Length 656;
Best Local Similarity 27.9%; Pred. No. 0.33;
Matches 48; Conservative 10; Mismatches 73; Indels 41; Gaps 9;

```

```

QY 29 YAGTVLLCOA-NOEGSSMTSGPSLVYTSAMGFPYPATATAAAYRGALHGRGRVTYNT 87
DB 29 YAGTT---QAYGQSQSYGTGQPTDVSYTOAQTATYGTOTAYATSY-----GQPTGYTT 79
QY 88 FRAAAPPPPIPAYGVVQEPYVGNKLLGGGYAAYRYAQPPTATAAAGDSYGRVYAADP 147
DB 80 -----PTAPQ---AYSQPVQGY-----GTGAYTTTATYTTTQA-----SYAAS 116
QY 148 YHHTLAPAPTYGVGAMNAFAPLTDKTRSHADVDGLVLSLQASIGGGYNR 199
DB 117 AYGTQPAVPAVQ-----QPAATAPTRPODGNKPTETSQPSS--TGGYNQ 160

```

```

RESULT 9
US-09-214-564A-4
; Sequence 4, Application US/09214564A
; Patent No. 6150515
; GENERAL INFORMATION:
; APPLICANT: Sharp, Phillip A.

```

```

; APPLICANT: Zhou, Qiang
; TITLE OF INVENTION: TAT-SF: Cofactor For Stimulation Of Transcriptional
; TITLE OF INVENTION: Elongation By HIV-1 TAT
; FILE REFERENCE: M0656/7042
; CURRENT APPLICATION NUMBER: US/09/214,564A
; CURRENT FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: US 60/021,218
; PRIOR FILING DATE: 1996-07-03
; PRIOR APPLICATION NUMBER: US 60/033,152
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: PCT/US97/11713
; PRIOR FILING DATE: 1997-07-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 656
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-214-564A-4

```

```

Query Match      8.5%; Score 92.5; DB 4; Length 656;
Best Local Similarity 27.9%; Pred. No. 0.33;
Matches 48; Conservative 10; Mismatches 73; Indels 41; Gaps 9;

```

```

QY 29 YAGTVLLCOA-NOEGSSMTSGPSLVYTSAMGFPYPATATAAAYRGALHGRGRVTYNT 87
DB 29 YAGTT---QAYGQSQSYGTGQPTDVSYTOAQTATYGTOTAYATSY-----GQPTGYTT 79
QY 88 FRAAAPPPPIPAYGVVQEPYVGNKLLGGGYAAYRYAQPPTATAAAGDSYGRVYAADP 147
DB 80 -----PTAPQ---AYSQPVQGY-----GTGAYTTTATYTTTQA-----SYAAS 116
QY 148 YHHTLAPAPTYGVGAMNAFAPLTDKTRSHADVDGLVLSLQASIGGGYNR 199
DB 117 AYGTQPAVPAVQ-----QPAATAPTRPODGNKPTETSQPSS--TGGYNQ 160

```

```

RESULT 10
US-09-343-011B-2
; Sequence 2, Application US/09343011B
; Patent No. 6300473
; GENERAL INFORMATION:
; APPLICANT: Stephane Richard
; TITLE OF INVENTION: SLM-1 AND SLM-2; NOVEL
; TITLE OF INVENTION: SAM68-LIKE MAMMALIAN PROTEINS
; FILE REFERENCE: A32561
; CURRENT APPLICATION NUMBER: US/09/343,011B
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: CA 2265271
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Mus musculus
; US-09-343-011B-2

```

```

Query Match      8.4%; Score 91; DB 4; Length 346;
Best Local Similarity 33.3%; Pred. No. 0.2;
Matches 32; Conservative 8; Mismatches 46; Indels 10; Gaps 4;

```

```

QY 60 GPPYPAATATAAAYRGALHGRG-----RTVYNT-FRAAAPPPPIPAYGVVQEPYVGN 112
DB 221 GPTTRGVST--RGPRVGRGLLTPRARGVPTGVRPPPTGTETGEYVDG-YGT 277
QY 113 KLLGGYAAVRYAQPPTATAAAYSDSYGRVYAADP 148
DB 278 AYDEQSYDSYDSYSTPAQASADYDYDGHGLEDAY 313

```

```

RESULT 11
US-09-442-100-8

```


Sequence 8, Application US/09442100
Patent No. 6359193
GENERAL INFORMATION:
APPLICANT: Xu, Tian
APPLICANT: Tao, Wufan
APPLICANT: Wang, Weiyl
APPLICANT: Zhang, Sheng
TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF LATS
TITLE OF INVENTION: GENES AND METHODS BASED THEREON
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/442,100
CLASSIFICATION:
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/411,111
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6523-003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 980 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-09-442-100-8

Query Match 8.1%; Score 87.5; DB 4; Length 980;
Best Local Similarity 25.4%; Pred. No. 1.8;
Matches 50; Conservative 21; Mismatches 81; Indels 45; Gaps 9;

QY 26 PDFAVGVLLQANQEGSSMTSPSLVYTSAMP-----GFPYPAATAAAYRGAHLRGRG 81
DB 108 PSEFGTGBALPSYHQLGGANGYEGPAL--EEMPROYIDFLPG--AGAGTHGAQHQP 162
QY 82 RTVNTFRRAAPRPPIR---AYGTVVQEPVYG-----NKLLQGVAAHYVAQ--- 126
DB 163 PKGST--AVEPSAHFPGTHGRHLLSEQSGYGVGRSSFNQNTPRPAYSSMKAQGGP 220
QY 127 ----PTPATAAYSDSYGRVYAADPYHTTLPAPTYG--VGAMNAFPLTDATRSHAD 179
DB 221 PASLTFFAHAGLVYASHHKPAATPGAHPLHVLGTRGPTFTGESGAQVLAPSR----- 274
QY 180 DVGVLSSLAQSIYQG 196
DB 275 -----NSLNDLYELG 285

RESULT 12
US-08-552-142A-11
Sequence 11, Application US/08552142A
Patent No. 5695995
GENERAL INFORMATION:
APPLICANT: Weintraub, Harold M.

APPLICANT: Lee, Jacqueline E.
APPLICANT: Tapscott, Stephen J.
APPLICANT: Hollenberg, Stanley M.
TITLE OF INVENTION: Neurogenic Differentiation (Neurod) Genes
TITLE OF INVENTION: and Proteins
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
STREET: 1420 Fifth Avenue, Suite 2800
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/552,142A
FILING DATE: 02-NOV-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,238
FILING DATE: 06-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US95/05741
FILING DATE: 08-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Broderick, Thomas F.
REGISTRATION NUMBER: 31,312
REFERENCE/DOCKET NUMBER: FPCR-1-8933
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-682-8100
TELEFAX: 206-225-0709
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 379 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-552-142A-11

Query Match 8.0%; Score 87; DB 1; Length 379;
Best Local Similarity 20.9%; Pred. No. 0.58;
Matches 38; Conservative 23; Mismatches 73; Indels 48; Gaps 6;

QY 14 WKLPVVGAVVSPDFVAGTVLLC-----QANQEGSSMTSGP 49
DB 171 WALSEILRSGRPRPLVYVQTLCKLSQPTTNLVAGCLQNSRNFLEQGRDGAHRRHGS 230
QY 50 SSLVYTSAMPGFYPYPA-----TAAAYRGAHLRGRG-TVYNT-FRAAADPPPIPA 99
DB 231 GG---PRAMHPYRYPGSRGRGTVPGAAMAAGARLRTHGCAVETLYAAAGGASPD 287
QY 100 YGVVYQEPVYGNLLOGGTAATRYAOPTPATAAAYSDSYGRVYAADPYHTTLPAPTYG 159
DB 288 YNSSEYEGPLSPPLCLNGNPSLKDSSPDHEKSYHS-----MHVSGCGSRNHG 336
QY 160 VG 161
DB 337 HG 338

RESULT 13
US-08-946-914-14
Sequence 14, Application US/08946914
Patent No. 6027916
GENERAL INFORMATION:
APPLICANT: Ni, Jian
APPLICANT: Gentz, Retner L.

APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESS: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,914
FILING DATE: Herewith
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,093
FILING DATE: 09-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGM
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 262 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-946-914-14

Query Match 8.0%; Score 86.5; DB 3; Length 262;
Best Local Similarity 26.1%; Pred. No. 0.4;
Matches 46; Conservative 7; Mismatches 46; Indels 77; Gaps 11;

QY 8 NPYTNGMKLNPVGAIVSPDFYAGTVLLCOANOEGSSMYSGPSSLYTSGAMP-----GF 61
DB 16 NPNPGWMP-----GA-----WGNQPGAGYPGAS---YPGAYPGQAPPGGY 53
QY 62 P-----YPAATAAAYRGALHGRGRTVYNTFRAAPRPPPIPAVGGVVQEPVYGNKL 114
DB 54 PGQAPPSAIVPGPTGPSAIVPGP-----TAPGAYPGPTAP---GAPPGQP----- 93

QY 115 LOGGYAAVRYAQPPTATAAAYSDSYGRVYADPYHHTLAPAPTYGVGMNAFAPLT 170
DB 94 --GGPGAY-----PSAPGAYPSAPGAYPATGPF-----GAPLT-----GPLT 127

RESULT 14
US-09-656-450-14
Sequence 14, Application US/09656450
Patent No. 6468768
GENERAL INFORMATION:
APPLICANT: NI, Jian
APPLICANT: Gentz, Retner L.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Galectin 9 and 10SV Polynucleotides
FILE REFERENCE: 1488.0560003
CURRENT APPLICATION NUMBER: US/09/656,450
CURRENT FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: US 09/263,689
PRIOR FILING DATE: 1999-03-05
PRIOR APPLICATION NUMBER: US 08/946,914
PRIOR FILING DATE: 1997-10-09
PRIOR APPLICATION NUMBER: US 60/028,093
PRIOR FILING DATE: 1996-10-09

NUMBER OF SEQ ID NOS: 60
SOFTWARE: Patentin version 3.0
SEQ ID NO: 14
LENGTH: 262
TYPE: PRT
ORGANISM: Rat
US-09-656-450-14

Query Match 8.0%; Score 86.5; DB 4; Length 262;
Best Local Similarity 26.1%; Pred. No. 0.4;
Matches 46; Conservative 7; Mismatches 46; Indels 77; Gaps 11;

QY 8 NPYTNGMKLNPVGAIVSPDFYAGTVLLCOANOEGSSMYSGPSSLYTSGAMP-----GF 61
DB 16 NPNPGWMP-----GA-----WGNQPGAGYPGAS---YPGAYPGQAPPGGY 53
QY 62 P-----YPAATAAAYRGALHGRGRTVYNTFRAAPRPPPIPAVGGVVQEPVYGNKL 114
DB 54 PGQAPPSAIVPGPTGPSAIVPGP-----TAPGAYPGPTAP---GAPPGQP----- 93

QY 115 LOGGYAAVRYAQPPTATAAAYSDSYGRVYADPYHHTLAPAPTYGVGMNAFAPLT 170
DB 94 --GGPGAY-----PSAPGAYPSAPGAYPATGPF-----GAPLT-----GPLT 127

RESULT 15
US-08-658-136-5
Sequence 5, Application US/08658136
Patent No. 6071717
GENERAL INFORMATION:
APPLICANT: KLINGER, KATHERINE W
APPLICANT: LANDES, GREGORY M
APPLICANT: BURN, TIMOTHY C
APPLICANT: CONNORS, TIMOTHY D
APPLICANT: DACKOWSKI, WILLIAM
APPLICANT: GERMINO, GREGORY
APPLICANT: QIAN, PENG
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: ONE MOUNTAIN ROAD
CITY: FRAMINGHAM
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/658,136
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LASSEN, ELIZABETH
REGISTRATION NUMBER: 31,845
REFERENCE/DOCKET NUMBER: GEN4-17.8
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-872-8400
TELEFAX: 508-872-5415
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4302 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-658-136-5

Query Match 8.0%; Score 86.5; DB 3; Length 4302;
Best Local Similarity 24.3%; Pred. No. 17;
Matches 55; Conservative 22; Mismatches 80; Indels 69; Gaps 10;

QY 26 PDFYAGTVLLCOANQEGSSMTSGPSSLVYTSAMPGFYPAA-----TAAAYRG 74
Db 185 PDNSSGTV-----AAVSFAAHEGLQPEACSAFCFSTGQGLAALSEQ 228
QY 75 AHLRGGRGVNTFR-----AAAPPPPIPAVGIVYQEPVY-----GNKLLOG 118
Db 229 WCLGGAOPSSASFACLSLCGPPPPPAFTCRGPTLLQHVFPASGATLVGPHGPLASGQ 288
QY 119 YAAVRYAOPTPATAA--YSDSYGRVYAADP-----YHHTLAPAPTYGVGMNA 165
Db 289 LAAFHIAAPLPVTATRWDFGDSAEVDAAGPAASHRYVLPGRYHTAVLA----LGAGSA 344
QY 166 FAPLTDKTRSHADVGIVL-----SLQASIV-QGYNRFAPY 203
Db 345 LIG-TDVQVEAAPALBLVCPSSVQSDSLDLSIQNRGSGLEBAAY 389

Search completed: March 13, 2003, 00:55:08
Job time : 20 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2003, 22:21:36 ; Search time 20 Seconds
(without alignments)
428.027 Million cell updates/sec

Title: US-09-809-545A-2

Perfect score: 1086
Sequence: 1 MTNKKAVNPYTNMGKLNPNVY.....VLSSLQASLYOGGYNRFAPY 203

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 188354 seqs, 42170167 residues

Total number of hits satisfying chosen parameters: 188354

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Published Applications_AA:*
2: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB pep:*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB pep:*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB pep:*
6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB pep:*
7: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB pep:*
11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB pep:*
12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB pep:*
13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB pep:*
14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1086	100.0	203	US-09-809-545A-2	Sequence 2, Appl1
2	551	50.7	330	US-09-794-591-2	Sequence 2, Appl1
3	96	8.8	760	US-08-754-311B-2	Sequence 2, Appl1
4	86.5	8.0	262	US-10-235-674-14	Sequence 14, Appl1
5	86.5	8.0	262	US-09-263-689-14	Sequence 14, Appl1
6	86	7.9	245	US-09-796-858-20	Sequence 20, Appl1
7	85	7.8	275	US-09-976-414-7	Sequence 7, Appl1
8	85	7.8	275	US-09-060-854B-3	Sequence 3, Appl1
9	85	7.8	275	US-10-075-895-2	Sequence 2, Appl1
10	85	7.8	275	US-10-075-895-2	Sequence 2, Appl1
11	85	7.7	382	US-10-050-624-31	Sequence 31, Appl1
12	83.5	7.7	485	US-09-925-300-1664	Sequence 1664, Ap
13	83	7.6	450	US-09-765-111A-14	Sequence 14, Appl1
14	82.5	7.6	400	US-09-073-009-126	Sequence 126, App
15	82.5	7.6	499	US-09-793-106-126	Sequence 126, App
16	82.5	7.6	499	US-09-863-877-2	Sequence 2, Appl1
17	82.5	7.6	499	US-09-732-350-2	Sequence 2, Appl1
18	81	7.5	551	US-09-920-300A-1789	Sequence 1789, Ap
19	81	7.5	551	US-10-033-528-1789	Sequence 1789, Ap

20	80.5	7.4	550	9	US-09-976-740-47	Sequence 47, Appl1
21	80.5	7.4	550	12	US-10-023-529-47	Sequence 47, Appl1
22	80.5	7.4	550	12	US-10-023-523-47	Sequence 47, Appl1
23	80	7.4	347	10	US-09-844-353A-100	Sequence 100, App
24	80	7.4	450	10	US-09-765-111A-25	Sequence 25, Appl1
25	80	7.4	527	9	US-09-712-163-156	Sequence 156, App
26	80	7.4	529	9	US-10-046-420-2	Sequence 2, Appl1
27	80	7.4	639	10	US-09-782-906-5	Sequence 5, Appl1
28	80	7.4	1062	10	US-09-815-242-5111	Sequence 511, Ap
29	79.5	7.3	171	10	US-09-925-297-659	Sequence 659, App
30	79	7.3	381	10	US-09-920-118-16	Sequence 16, Appl1
31	79	7.3	723	12	US-10-044-205A-32	Sequence 32, Appl1
32	79	7.3	726	12	US-10-044-205A-31	Sequence 31, Appl1
33	79	7.3	3788	9	US-10-014-717-6	Sequence 6, Appl1
34	78.5	7.2	465	9	US-10-092-063-39	Sequence 39, Appl1
35	78	7.2	275	10	US-09-837-235-17	Sequence 17, Appl1
36	78	7.2	275	10	US-09-060-854B-4	Sequence 4, Appl1
37	78	7.2	302	9	US-10-029-180-18	Sequence 18, Appl1
38	78	7.2	357	10	US-09-837-235-15	Sequence 15, Appl1
39	77.5	7.1	240	10	US-09-819-917-8	Sequence 8, Appl1
40	77.5	7.1	382	10	US-09-809-033A-3	Sequence 3, Appl1
41	77.5	7.1	382	10	US-09-809-033A-4	Sequence 4, Appl1
42	77.5	7.1	399	9	US-10-029-180-64	Sequence 64, Appl1
43	77.5	7.1	1094	9	US-09-712-363-287	Sequence 287, App
44	77	7.1	265	10	US-09-819-254-2	Sequence 2, Appl1
45	77	7.1	311	9	US-10-025-380-1061	Sequence 1061, Ap

ALIGNMENTS

```
RESULT 1
US-09-809-545A-2
: Sequence 2, Application US/09809545A
: Patent No. US20020110804A1
: GENERAL INFORMATION:
: APPLICANT: Stanton, Lawrence W.
: APPLICANT: White, R. Tyler
: TITLE OF INVENTION: SECRETED FACTORS
: FILE REFERENCE: SCTOS.017A
: CURRENT APPLICATION NUMBER: US/09/809,545A
: CURRENT FILING DATE: 2001-03-14
: NUMBER OF SEQ ID NOS: 84
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 203
: TYPE: PRT
: ORGANISM: Rattus norvegicus
US-09-809-545A-2

Query Match      100.0%; Score 1086; DB 10; Length 203;
Best Local Similarity 100.0%; Pred. No. 1,3e-94;
Matches 203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTNKKAVNPYTNMGKLNPNVGVAVYSPDFYACTVLLCQANQCGSSMYSGPSSLVYTSAMPG 60
DB 1 MTNKKAVNPYTNMGKLNPNVGVAVYSPDFYACTVLLCQANQCGSSMYSGPSSLVYTSAMPG 60
QY 61 PPYPATTAAYRGAHLRGRGRYTYNTFRRAAPPPIPAVGYGVYQEPVYGNKLLQGGYA 120
DB 61 PPYPATTAAYRGAHLRGRGRYTYNTFRRAAPPPIPAVGYGVYQEPVYGNKLLQGGYA 120
QY 121 AYRYQPTPATTAAYSDSGRYVAADPYHHTLAPATYGVGAMNNAFAPLTDKTRSHADD 180
DB 121 AYRYQPTPATTAAYSDSGRYVAADPYHHTLAPATYGVGAMNNAFAPLTDKTRSHADD 180
QY 181 VGLVLSLQASLYOGGYNRFAPY 203
DB 181 VGLVLSLQASLYOGGYNRFAPY 203

RESULT 2
US-09-794-591-2
```

; Sequence 2, Application US/09794591
; Patent No. US20010018198A1
; GENERAL INFORMATION:
; APPLICANT: Shiba, Stefan M.
; APPLICANT: Shiba, Hiroki
; TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
; TITLE OF INVENTION: Products Related Thereto and Methods of Using Same
; FILE REFERENCE: CE 3093
; CURRENT APPLICATION NUMBER: US/09/794,591
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/145,391
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 2
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-794-591-2

Query Match 50.7%; Score 551; DB 10; Length 330;
Best Local Similarity 97.1%; Pred. No. 3, 4e-44;
Matches 102; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MTNKAANDYNTGKMLNPNVGAIVSPDYAGTVLLCOANQEGSSMYSGPSLLVYTSAMFG 60
Db 196 MTNKTIVNTYTGKMLNPNVGAIVSPDYAGTVLLCOANQEGSSMYSAPSLLVYTSAMFG 255
Qy 61 PPYPATAAARGAHLRGRGRTVNTFPRAAPPPPIPAVGGVY 105
Db 256 PPYPATAAARGAHLRGRGRTVNTFPRAAPPPPIPAVGGVY 300

RESULT 3

US-08-754-311B-2
; Sequence 2, Application US/08754311B
; Patent No. US2002004221A1
; GENERAL INFORMATION:
; APPLICANT: Bonini, Nancy M.
; APPLICANT: Leiserson, William M.
; APPLICANT: Benzer, Seymour
; TITLE OF INVENTION: PROGRAMMED CELL DEATH ANTAGONIST
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobach, Test, Albrighton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/754,311B
; FILING DATE: 21-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/195,152
; FILING DATE: 14-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard F.
; REGISTRATION NUMBER: 31,801
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 760 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-754-311B-2

Query Match 8.8%; Score 96; DB 8; Length 760;
Best Local Similarity 23.7%; Pred. No. 0.44;
Matches 42; Conservative 24; Mismatches 61; Indels 50; Gaps 6;

Qy 42 GSSMY-----SGPSLLVYTSAMPPEPPYPAATAAARGAHLRGRGRTVNTFPRAAA 92
Db 223 GSNLYGSSASNPDLGAVANSSAV-----AAAAAVYDGKH---DYIYNSMOQYT 272
Qy 93 PPPPIPAVGGVYQEPYVGNKLLQGYAARYAQPPTATAA----- 134
Db 273 PPPFVSGVTPTAATAARQAAMERGAATAAAAYLTPTSAASGNNSQULYSSPYAGYNNF 332
Qy 135 -----YSDSYGRVYAADPYHHTLAPAP---TYGVG-AMNAFAPLTDATRSHA 178
Db 333 GQDDYGVYNEQYGVNYPANYSPYAVSSPSSSASHGHGFVAAASNLSESPDTDHS 389

RESULT 4

US-10-235-674-14
; Sequence 14, Application US/10235674
; Publication No. US20030040081A1
; GENERAL INFORMATION:
; APPLICANT: Ni, Jian
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Galactin 9 and 10SV Polynucleotides
; FILE REFERENCE: 1488.0560004
; CURRENT APPLICATION NUMBER: US/10/235,674
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: US 09/656,450
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 09/263,689
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: US 08/946,914
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: US 60/028,093
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Patent version 3.0
; SEQ ID NO 14
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Rat
US-10-235-674-14

Query Match 8.0%; Score 86.5; DB 9; Length 262;
Best Local Similarity 26.1%; Pred. No. 0.9;
Matches 46; Conservative 7; Mismatches 46; Indels 77; Gaps 11;

Qy 8 NPYTNGKMLNPNVGAIVSPDYAGTVLLCOANQEGSSMYSGPSLLVYTSAMP-----GF 61
Db 16 NNPNQGW-----GA-----WGNQPGAGGYPGAS---YCGAYGQAAPGQY 53
Qy 62 P-----YPAATAAARGAHLRGRGRTVNTFPRAAPPPPIPAVGGVYQEPYVGNKLL 114
Db 54 PQQAPPSAVPGPTGSAVPGP-----TABGAYFGPAP--GAFPGQP----- 93
Qy 115 LQGYAARYAQAQPTATAAAYSDSYGRVYAADPYHHTLAPATYGVGAMNAPAPLT 170
Db 94 --GGPGAY-----PSAPGAYPSAPGAYPATGPF-----GAPT-----GPLT 127

RESULT 5

US-09-263-689-14
; Sequence 14, Application US/09263689
; Patent No. US20020150970A1
; GENERAL INFORMATION:

APPLICANT: Ni, Jian
APPLICANT: Gentz, Reiner L.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/263,689
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/946,914
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGM
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 262 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-263-689-14

Query Match
Best Local Similarity 26.1%; Score 86.5; DB 10; Length 252;
Matches 46; Conservative 7; Mismatches 46; Indels 77; Gaps 11;

QY 8 NPYNGMKLNIVGAVSPDFYAGVLLCQANOGSSMYSGPSSLVYTSAMP-----GF 61
DB 16 NPNQGNP-----GA-----WGNQPGAGCPGAS--YGAIFGQAPPGCY 53
QY 62 P-----YPATTAAYRGALHGRGRTVNTFRAPAPPPPIYAGGVVYQEPYGNKL 114
DB 54 PGQAPSAVPGPTGSAVPGP-----TAPGAYFGPTAP---GAPFGP----- 93
QY 115 LOGGYAAYRYAQPPTATTAAYSDSYGRVYAADPHYHTLAPPTYGVGAMNFAFLT 170
DB 94 --GGPGAY-----PSAPGAYPSAPGAYPATGPF-----GAPLT-----GFLT 127

RESULT 6
US-09-796-858-20
Sequence 20, Application US/09796858
Patent No. US20020055139A1
GENERAL INFORMATION:
APPLICANT: Holtzmann, Douglas
TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING PROGNOSTIC, DIAGNOSTIC,
FILE REFERENCE: 7853-226-999
CURRENT APPLICATION NUMBER: US/09/796,858
PRIOR FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 09/223,094
PRIOR FILING DATE: 1998-12-30
PRIOR APPLICATION NUMBER: 09/223,546
PRIOR FILING DATE: 1998-12-30
PRIOR APPLICATION NUMBER: 09/224,246

PRIOR FILING DATE: 1998-12-30
PRIOR APPLICATION NUMBER: 09/312,359
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 09/336,536
PRIOR FILING DATE: 1999-06-18
PRIOR APPLICATION NUMBER: 09/342,687
PRIOR FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: 09/399,723
PRIOR FILING DATE: 1999-09-20
PRIOR APPLICATION NUMBER: 09/471,179
PRIOR FILING DATE: 1999-12-23
PRIOR APPLICATION NUMBER: 09/474,071
PRIOR FILING DATE: 1999-12-29
PRIOR APPLICATION NUMBER: 09/474,072
PRIOR FILING DATE: 1999-12-29
PRIOR APPLICATION NUMBER: 09/572,002
PRIOR FILING DATE: 2000-05-14
PRIOR APPLICATION NUMBER: 09/597,993
PRIOR FILING DATE: 2000-06-12
PRIOR APPLICATION NUMBER: 09/599,596
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 09/606,565
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: 09/365,164
PRIOR FILING DATE: 1999-07-30
PRIOR APPLICATION NUMBER: 09/630,334
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: 09/665,666
PRIOR FILING DATE: 2000-09-20
NUMBER OF SEQ ID NOS: 50
SEQ ID NO 20
LENGTH: 245
TYPE: PRT
ORGANISM: Homo sapiens
US-09-796-858-20

Query Match
Best Local Similarity 7.9%; Score 86; DB 10; Length 245;
Matches 52; Conservative 9; Mismatches 51; Indels 100; Gaps 11;

QY 26 PDFYAGT-----VLLCQANQF-----GSSMYSG-----PSSLVYTSA----- 57
DB 55 PDFCCGSSQYCCSDVLKXIOWNEWCPEPSSRSFSAHPETPOLQSALUKYOSLSDSN 114
QY 58 MPGRPYATTAAYRGALHGRGRTVNTFRAP-----GLTFVVFPIATIIYVCTSCCCCLYKMCRRPVPVSN 91
DB 115 MPGR-----GATYAT-----GLTFVVFPIATIIYVCTSCCCCLYKMCRRPVPVSN 161
QY 92 -----APPPIYAGGVVYQEPYGNKLLGGYAAVRYAQPPTA-TAAAYSDSY 139
DB 162 TTTVHTAVYQGPVAVSFGPTFYQ-----GYHMPPOPGMFAAPFYPTQY 207
QY 140 GRVYAADP-----YHHTLAPA--PTYGVGAMN 164
DB 208 PPPLYAQPPTGPPAYHETLAGASQPPVNPAYMD 239

RESULT 7
US-09-976-414-7
Sequence 7, Application US/09976414
Patent No. US2002015575A1
GENERAL INFORMATION:
APPLICANT: No. US2002015575A1regard-Madsen, Made
APPLICANT: Larsen, Line
APPLICANT: Hansen, Peter
TITLE OF INVENTION: Subtilase Variants
FILE REFERENCE: 10081.200-US
CURRENT APPLICATION NUMBER: US/09/976,414
CURRENT FILING DATE: 2001-10-12
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7
LENGTH: 275

TYPE: PRT
ORGANISM: Bacillus
US-09-976-414-7

Query Match 7.8%; Score 85; DB 9; Length 275;
Best Local Similarity 26.4%; Pred. No. 1.3;
Matches 46; Conservative 18; Mismatches 82; Indels 28; Gaps 7;

QY 3 NKKAVNPYNGKLNIPVGAAYSPDFYAGTVLLCOANOGSSMYSGPSLVYTSAMPGP 62
DB 118 NMDVIMSLGSGSSGSAALKAADVKAASGVVVAAGNCGT---SGSSSTV-----GYP 168
QY 63 --YPAATAAAAYRGALHGRGRVTYNTFRAAAPPPPIPAYGGVYQEPYGNKLLGGYA 120
DB 169 GKYPGVIAVGAVDSSNQRA-----SFSSVGPBLDVA--PGVSIQSTLPGNK--YGAYN 218
QY 121 AYRYQPTPTATAAYSDSYGRVYAADPYHHTLAPPT-----YGVGMNAFA 167
DB 219 GTSMASPHVGAALILSKHPMWTNTQVRSSLENTTKLGDSFYGKGLINVOA 272

RESULT 8
US-09-060-854B-3
Sequence 3, Application US/09060854B
Patent No. US2002081703A1
GENERAL INFORMATION:
APPLICANT: Esceil, David Aaron
TITLE OF INVENTION: Human Protease and Use of Such Protease for Pharmaceutical
TITLE OF INVENTION: Applications and for Reducing the Allergenicity of No. US2002008
FILE REFERENCE: GC532
CURRENT APPLICATION NUMBER: US/09/060,854B
CURRENT FILING DATE: 1998-04-15
NUMBER OF SEQ ID NOS: 7
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 275
TYPE: PRT
ORGANISM: B. amyloliquefaciens
US-09-060-854B-3

Query Match 7.8%; Score 85; DB 10; Length 275;
Best Local Similarity 26.4%; Pred. No. 1.3;
Matches 46; Conservative 18; Mismatches 82; Indels 28; Gaps 7;

QY 3 NKKAVNPYNGKLNIPVGAAYSPDFYAGTVLLCOANOGSSMYSGPSLVYTSAMPGP 62
DB 118 NMDVIMSLGSGSSGSAALKAADVKAASGVVVAAGNCGT---SGSSSTV-----GYP 168
QY 63 --YPAATAAAAYRGALHGRGRVTYNTFRAAAPPPPIPAYGGVYQEPYGNKLLGGYA 120
DB 169 GKYPGVIAVGAVDSSNQRA-----SFSSVGPBLDVA--PGVSIQSTLPGNK--YGAYN 218
QY 121 AYRYQPTPTATAAYSDSYGRVYAADPYHHTLAPPT-----YGVGMNAFA 167
DB 219 GTSMASPHVGAALILSKHPMWTNTQVRSSLENTTKLGDSFYGKGLINVOA 272

RESULT 9
US-10-075-907-2
Sequence 2, Application US/10075907
Patent No. US20020127695A1
GENERAL INFORMATION:
APPLICANT: Genecor International, Inc.
APPLICANT: Davis, Benjamin G.
APPLICANT: Jones, Bryan J.
APPLICANT: Bott, Richard R.
TITLE OF INVENTION: CHEMICALLY MODIFIED ENZYMES WITH
FILE REFERENCE: 23623-7039
CURRENT APPLICATION NUMBER: US/10/075,907
CURRENT FILING DATE: 2002-02-12
PRIOR APPLICATION NUMBER: US 60/113,130

PRIOR FILING DATE: 1998-12-21
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 275
TYPE: PRT
ORGANISM: Bacillus amyloliquefaciens
FEATURE:
OTHER INFORMATION: Subtilisin BPN'
US-10-075-907-2

Query Match 7.8%; Score 85; DB 12; Length 275;
Best Local Similarity 26.4%; Pred. No. 1.3;
Matches 46; Conservative 18; Mismatches 82; Indels 28; Gaps 7;

QY 3 NKKAVNPYNGKLNIPVGAAYSPDFYAGTVLLCOANOGSSMYSGPSLVYTSAMPGP 62
DB 118 NMDVIMSLGSGSSGSAALKAADVKAASGVVVAAGNCGT---SGSSSTV-----GYP 168
QY 63 --YPAATAAAAYRGALHGRGRVTYNTFRAAAPPPPIPAYGGVYQEPYGNKLLGGYA 120
DB 169 GKYPGVIAVGAVDSSNQRA-----SFSSVGPBLDVA--PGVSIQSTLPGNK--YGAYN 218
QY 121 AYRYQPTPTATAAYSDSYGRVYAADPYHHTLAPPT-----YGVGMNAFA 167
DB 219 GTSMASPHVGAALILSKHPMWTNTQVRSSLENTTKLGDSFYGKGLINVOA 272

RESULT 10
US-10-075-895-2
Sequence 2, Application US/10075895
Patent No. US2002013717A1
GENERAL INFORMATION:
APPLICANT: Genecor International, Inc.
APPLICANT: Jones, Bryan J.
TITLE OF INVENTION: MODIFIED ENZYMES AND THEIR USE FOR
FILE REFERENCE: 23623-7013
CURRENT APPLICATION NUMBER: US/10/075,895
CURRENT FILING DATE: 2002-02-13
PRIOR APPLICATION NUMBER: US 60/072,265
PRIOR FILING DATE: 1998-01-23
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 275
TYPE: PRT
ORGANISM: Bacillus amyloliquefaciens
FEATURE:
OTHER INFORMATION: Subtilisin BPN'
US-10-075-895-2

Query Match 7.8%; Score 85; DB 12; Length 275;
Best Local Similarity 26.4%; Pred. No. 1.3;
Matches 46; Conservative 18; Mismatches 82; Indels 28; Gaps 7;

QY 3 NKKAVNPYNGKLNIPVGAAYSPDFYAGTVLLCOANOGSSMYSGPSLVYTSAMPGP 62
DB 118 NMDVIMSLGSGSSGSAALKAADVKAASGVVVAAGNCGT---SGSSSTV-----GYP 168
QY 63 --YPAATAAAAYRGALHGRGRVTYNTFRAAAPPPPIPAYGGVYQEPYGNKLLGGYA 120
DB 169 GKYPGVIAVGAVDSSNQRA-----SFSSVGPBLDVA--PGVSIQSTLPGNK--YGAYN 218
QY 121 AYRYQPTPTATAAYSDSYGRVYAADPYHHTLAPPT-----YGVGMNAFA 167
DB 219 GTSMASPHVGAALILSKHPMWTNTQVRSSLENTTKLGDSFYGKGLINVOA 272

RESULT 11
US-10-090-624-31
Sequence 31, Application US/10090624
Patent No. US20020132335A1


```

GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: SHIMOJO, Tomoko
APPLICANT: ASADA, Kiyozo
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/10/090,624
CURRENT FILING DATE: 2002-03-06
PRIOR APPLICATION NUMBER: 09/445,472
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
PRIOR FILING DATE: 1997-06-10
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn version 3.0
SEQ ID NO 31
LENGTH: 382
TYPE: PRT
ORGANISM: Bacillus subtilis
US-10-090-624-31

```

```

Query Match 7.8%; Score 85; DB 12; Length 382;
Best Local Similarity 26.4%; Pred. No. 2;
Matches 46; Conservative 18; Mismatches 82; Indels 28; Gaps 7;

```

```

QY 3 NKKAVNPPTNGMKLNPPVAVSPDYAGTVLLCOANQSGSSMYSGPSLVTYSAMPGF 62
DB 225 NMDVYNNLSGSPSSAALKAVDCAVAGVYVAAAGNGEGR---SGSSSTV-----GYF 275

QY 63 --YPAATAAAYRGAAHLRGRGRVYNTFRAPAPPPIPAYGVVQEPVYGNKLLGGYA 120
DB 276 GKYSVLAIVAGVDSNQA-----SPSSVGPFLDVMA-PGVSIQSTLPGNK--YGAYN 325

QY 121 AYRAQPTPAATAAAYSDSYGRVYAADPYHHTLAPPT-----YGVANMAFA 167
DB 326 GTSMASPHVAGAAALILSKHPMTNTQVRSSELENTTKLGDSPYKGLINVOA 379

```

```

RESULT 12
US-09-925-300-1664
Sequence 1664, Application US/09925300
Patent No. US20020151681A1
GENERAL INFORMATION:
APPLICANT: Craig Rosen,
APPLICANT: Steve Ruben
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA101
CURRENT APPLICATION NUMBER: US/09/925,300
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05988
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1890
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1664
LENGTH: 485
TYPE: PRT
ORGANISM: Homo sapiens
US-09-925-300-1664

```

```

Query Match 7.7%; Score 83.5; DB 10; Length 485;
Best Local Similarity 24.8%; Pred. No. 3.7;
Matches 39; Conservative 15; Mismatches 70; Indels 33; Gaps 8;

```

```

QY 9 PYTGMKLNPPVAVSPDYAGTVLLCOANQSGSSMYSGPSLVTYSAMPGFYPAATA 68
DB 52 PYPGGRF--PWGGAGYPP-----VSSGYPGAGGYPAGCGYAPAGCGYGAQPG--G 99

QY 69 AAAYRGAAHL-RGRGRVYNTFRAPAPPPIPAYGVVQEPVYGNKLLGGYAARY--A 125
DB 100 APSYVGPVPGCGFVPPGAGFSGYPPPSQSYGGGPAQVPLPGG--PPGGQMSQYPPG 157

```

```

QY 126 QPT-----PATMAAYSDSYGRVYAADPYHHTLAPPT 158
DB 158 QPTYSQPATVTVTQTG-----TIRPANF 182

```

```

RESULT 13
US-09-765-111A-14
Sequence 14, Application US/09765111A
Patent No. US20020106796A1
GENERAL INFORMATION:
APPLICANT: Fletcher, Jonathan A.
APPLICANT: Kroll, Todd G.
TITLE OF INVENTION: PAX8-PARAGAMMA NUCLEIC ACID MOLECULES
FILE REFERENCE: B0801/7196/ERP/MAT
CURRENT APPLICATION NUMBER: US/09/765,111A
CURRENT FILING DATE: 2001-01-18
PRIOR APPLICATION NUMBER: US 60/177,109
PRIOR FILING DATE: 2000-01-20
PRIOR APPLICATION NUMBER: US 60/225,079
PRIOR FILING DATE: 2000-08-14
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 14
LENGTH: 450
TYPE: PRT
ORGANISM: Homo Sapiens
US-09-765-111A-14

```

```

Query Match 7.6%; Score 83; DB 10; Length 450;
Best Local Similarity 26.5%; Pred. No. 3.8;
Matches 49; Conservative 12; Mismatches 68; Indels 56; Gaps 8;

```

```

QY 18 PVGAVVSPDYAGTVLLCOANQSGSSMYSGPSLVT-----SAMP---GFTYPA 65
DB 296 PVVADPHSP-----FAIKETPEVSSSSSTPSSLSAFLDLOQVSGVPPFPAFAA 349

QY 66 ATAAAYRGAAHLRGR--GRVYNTFRAPAPPPIPAYGVVQEPVYGNKLLGGYAAY 122
DB 350 SVYQFTGQALLSRENVGPTL-----PGYPRHPIPTSG-----QGSYAS- 388

QY 123 RYAGPTPAATAAAYSDSYGRVYAADPY-----HHTLAPPTVYGANAPAPLTA 172
DB 389 ---SALAGWAGSEYSGNAVGHPTPSSSEAMRFNSSLSPYYSSTSRPAPPTA 444

QY 173 KTRSH 177
DB 445 TAFDH 449

```

```

RESULT 14
US-09-073-009-126
Sequence 126, Application US/09073009
Patent No. US20010012888A1
GENERAL INFORMATION:
APPLICANT: Alderson, Mark
APPLICANT: Dillon, Davin C.
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Campos-Neto, Antonio
TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
NUMBER OF SEQUENCES: 144
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Ave.
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

TUBERCULOSIS AND MEJ

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/073.009
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.441C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-632-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 126:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-073-009-126

Query Match 7.6%; Score 82.5; DB 10; Length 400;
Best Local Similarity 26.1%; Pred. No. 3.6;
Matches 48; Conservative 18; Mismatches 81; Indels 37; Gaps 8;

QY 34 LLQANQEGSSMYSGPSSLVYTSAMPFPYPATATAAAYRGALRGRTVYNTFRPA-- 91
DB 48 VTQLSSBG---WLGPAASAAAEVA--PYVAMMSAAQAQEAATQRAAAAAFEAAFA 102
QY 92 --APPPIPA-----YG---GVYQEPVYGNKLLGGYAAVRYAQTPTPATA 132
DB 103 ATVPPLIAANRASIMQLISTNVFGQNTSALAAEAQYGEWMAQDSAAAYAVAGSSASAS 162
QY 133 AAYSDSYGRVYAADPYHHTLAPARTYGVGAMNAPFLTDKTRSHADVGL--VLSSLOA 190
DB 163 AVTPSTPPQIANPTAQCTQAATAVTAAG--TAOSTLTEMIT-----GLPNALQSILTS 213
QY 191 SIYQ 194
DB 214 XLQ 217

RESULT 15
US-09-793-306-126
Sequence 126, Application US/09793306
Patent No. US20020098200A1
GENERAL INFORMATION:
APPLICANT: Campos-Neto, Antonio
APPLICANT: Skeiky, Yaelir
APPLICANT: Ovendale, Pamela
APPLICANT: Jen, Shylan
APPLICANT: Lodes, Michael
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
FILE REFERENCE: 014058-008740US
CURRENT APPLICATION NUMBER: US/09/793,306
CURRENT FILING DATE: 2001-02-26
PRIOR APPLICATION NUMBER: US 60/185,037
PRIOR FILING DATE: 2000-02-25
PRIOR APPLICATION NUMBER: US 60/223,828
PRIOR FILING DATE: 2000-08-08
NUMBER OF SEQ ID NOS: 164
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 126
LENGTH: 400
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: Tb470 (Mtb-40)
NAME/KEY: MOD_RES
LOCATION: (1)..(400)
OTHER INFORMATION: Xaa = any amino acid

US-09-793-306-126
Query Match 7.6%; Score 82.5; DB 10; Length 400;
Best Local Similarity 26.1%; Pred. No. 3.6;
Matches 48; Conservative 18; Mismatches 81; Indels 37; Gaps 8;

QY 34 LLQANQEGSSMYSGPSSLVYTSAMPFPYPATATAAAYRGALRGRTVYNTFRPA-- 91
DB 48 VTQLSSBG---WLGPAASAAAEVA--PYVAMMSAAQAQEAATQRAAAAAFEAAFA 102
QY 92 --APPPIPA-----YG---GVYQEPVYGNKLLGGYAAVRYAQTPTPATA 132
DB 103 ATVPPLIAANRASIMQLISTNVFGQNTSALAAEAQYGEWMAQDSAAAYAVAGSSASAS 162
QY 133 AAYSDSYGRVYAADPYHHTLAPARTYGVGAMNAPFLTDKTRSHADVGL--VLSSLOA 190
DB 163 AVTPSTPPQIANPTAQCTQAATAVTAAG--TAOSTLTEMIT-----GLPNALQSILTS 213
QY 191 SIYQ 194
DB 214 XLQ 217

Search completed: March 13, 2003, 00:53:19
Job time : 21 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_v2n model

Run on: March 16, 2003, 03:31:48 ; Search time 1706 Seconds

(without alignments)
2524.745 Million cell updates/sec

Title: US-09-809-545a-2_COPY_1_148
Sequence: 1 MTKKAVNPYTNCKMLNPV.....PATTAAYSDSYGRVADPY 148

Scoring table:

BLIGO
Xgapop 60.0 , Xgapext 60.0
Xgapop 60.0 , Xgapext 60.0
Xgapop 6.0 , Xgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4106490

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPNO.spool/US09809545/runat_07032003_153954_12159/app_query.fasta_1.327
-DB=genmb1 -QPM=fastcap -SUFFIX=olip2n.rge -MINMATCH=0.1 -LOOFC=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USR=US09809545 @CGN 1 1 3745 @runat_07032003_153954_12159 -NCPU=6 -ICPU=3
-NO_XLPPY -NO_MMAT -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Genembl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rtd:*
36: em_hcg_man:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	148	100.0	1340	6 AX268800	AX268800 Sequence
2	100	67.6	1363	10 AF107204	AF107204 Mus muscu
3	100	67.6	1547	9 AF094849	AF094849 Homo sapi
4	100	67.6	2000	9 AF229057	AF229057 Homo sapi
5	100	67.6	2002	10 AB041596	AB041596 Mus muscu
6	89	60.1	3348	9 AF109106	AF109106 Homo sapi
7	60	40.5	1586	10 AF191501	AF191501 Mus muscu
8	58	39.2	1475	9 AB060859	AB060859 Macaca fa
9	58	39.2	1513	9 AK001027	AK001027 Homo sapi
10	58	39.2	2279	9 AF107203	AF107203 Homo sapi
11	57	38.5	2372	6 AF134676	AF134676 Sequence
12	44	29.7	251	9 HSF0X09	AF109115 Homo sapi
13	44	29.7	164538	9 AC005774	AC005774 Homo sapi
14	34	23.0	596	9 HSA33368	AF094849 Homo sapi
15	34	23.0	617	9 HSA335811	AF229057 Homo sapi
16	34	23.0	653	9 HSA335616	AB041596 Mus muscu
17	34	23.0	675	9 HSA334013	AF109106 Homo sapi
18	21	14.2	151	9 HSF0X11	AF191501 Mus muscu
19	21	14.2	115860	2 AC120661	AB060859 Macaca fa
20	20	13.5	1623	9 HSM803066	AK001027 Homo sapi
21	17	11.5	374	9 HSF0X08	AF107203 Homo sapi
22	16	10.8	959	10 AF229056	AF134676 Sequence
23	14	9.5	249	9 HSF0X10	AF109115 Homo sapi
24	14	9.5	347	9 HSF0X07	AC005774 Homo sapi
25	14	9.5	112976	9 AC027683	AF094849 Homo sapi
26	14	9.5	115860	2 AC120661	AB060859 Macaca fa
27	13	8.8	1134	10 AF387322	AF229057 Homo sapi
28	13	8.8	1393	9 AY072786	AY072786 Homo sapi
29	13	8.8	1538	9 BC013115	BC013115 Homo sapi
30	13	8.8	1553	9 BC025281	BC025281 Homo sapi
31	13	8.8	1558	9 AK055213	AK055213 Homo sapi
32	13	8.8	1695	10 AF229055	AF229055 Mus muscu
33	13	8.8	1721	10 BC027263	BC027263 Mus muscu
34	13	8.8	1876	9 HSP18A20	AL009266 H. sapien
35	13	8.8	1925	9 AF228058	AF228058 Homo sapi
36	13	8.8	108315	9 HS41R2	AL049748 Human DNA
37	13	8.8	156549	2 AC079072	AC079072 Homo sapi
38	13	8.8	156820	2 AC123391	AC123391 Rattus no
39	13	8.8	159565	10 AL591075	AL591075 Mouse DNA
40	13	8.8	203649	10 AL603843	AL603843 Mouse DNA
41	13	8.8	208632	2 AC073624	AC073624 Homo sapi
42	13	8.8	209461	6 AC079359	AC079359 Homo sapi
43	12	8.1	428	6 AX336745	AX336745 Sequence
44	12	8.1	444	9 HSA33571	AJ343571 Homo sapi
45	12	8.1	669	9 HSA338965	AJ338965 Homo sapi

RESULT 1

ALIGNMENTS

AX268800
LOCUS AX268800 1340 bp DNA linear PAT 29-OCT-2001
DEFINITION Sequence 1 from Patent WO0174901.
ACCESSION AX268800
VERSION AX268800.1 GI:16541860
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 Stanton, L.W. and White, R.T.
TITLE Secreted factors
JOURNAL Patent: WO 0174901-A 1 11-OCT-2001;
Scienc Inc. (US)
FEATURES
source 1.1340
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
BASE COUNT 366 a 394 c 317 g 263 t
ORIGIN
Alignment Scores:
Pred. No.: 8,28e-119 Length: 1340
Score: 148.00 Matches: 148
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-809-545a-2_copy_1_148 (1-148) x AX268800 (1-1340)
QY 1 MetThrAsnLysLeuValAlaValAsnProTyrThrAsnGlyTyrPheLeuAsnProValAla 20
DB 535 ATGACTAAATTAAGAGCCCGTGAACCCCTACACCAATGCTGGAAATTAATTCAGATTG 54
QY 21 GAlaValAlaTyrSerProAspPheTyrAlaGlyThrValLeuLeuGlyGlnAlaAsnGln 40
DB 595 GCGCGCGTCTACAGCCCCGACCTCTATGACGACGCGTGTGTGTGCGCAGGCCAACAG 654
QY 41 GAlaGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
DB 655 GAGGAGATCTTCCATGATGACAGTGGCCCGACCTGCTGTATATACCTTGCATGCTGGC 714
QY 61 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyValAlaHisLeuArgGlyArg 80
DB 715 TTTCCATATCCGCGCCCGACCTGCTGACAGCTGATACCGAGGGCTACCTTCAGGCCGT 774
QY 81 GAlaArgThrValTyrAsnThrPheArgAlaAlaAlaProProProProAlaTyr 100
DB 775 GGTGCGACCGTGTACAAACCTTCAGAGCTGCGCGCCCCACCCCAATCCCGGCTAT 834
QY 101 GAlaGlyValAlaTyrGlnGluProValTyrGlyAsnLeuLeuGlnGlyTyrAla 120
DB 835 GCGCGAGTGTATCAAGAGCGCATGTATGGCAATTAATTCGTAAGGGTGTACGCT 894
QY 121 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSerTyrGly 140
DB 895 GCATACCGCTACGCCAGCCACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 954
QY 141 ArgValTyrAlaAlaAspProTyr 148
DB 955 CGAGTTATGCTGCCGACCCCTAC 978
RESULT 2
AF107204 1363 bp mRNA linear ROD 23-JUN-2000
LOCUS AF107204
DEFINITION Mus musculus ataxin 2-binding protein (A2bp) mRNA, complete cds.
ACCESSION AF107204
VERSION AF107204.1 GI:8671587
KEYWORDS
SOURCE Mus musculus.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 1363)
AUTHORS Shibata, H., Huynh, D.P., Vo, T.T. and Pulet, S.-M.
TITLE A novel protein, A2BP, with RNA binding motif, binds to C-terminal
ataxin-2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1363)
AUTHORS Shibata, H., Huynh, D.P., Vo, T.T. and Pulet, S.-M.
TITLE Direct Submision
JOURNAL Submitted (16-NOV-1998) Division of Neurology and Neurogenetics
Laboratory, Burns and Allen Research Institute, Cedars-Sinai
Medical Center, UCLA School of Medicine, 8700 Beverly Blvd., Los
Angeles, CA 90048, USA
FEATURES
source 1.1363
/organism="Mus musculus"
/db_xref="taxon:10090"
gene 1.1363
/gene="A2bp"
CDS 67..1203
/gene="A2bp"
/codon_start=1
/product="ataxin 2-binding protein"
/protein_id="AA078292.1"
/db_xref="GI:8671588"
/translation="MAQPYASQAPAPONGIAPAEYTAHPHAPAEYTGQTTPDHTLN
LYPTQHSBOSADTSAGTASGTATGTDAAPTGQPTQSENTESQOKRLHVN
IPRRFRPDPLROMFOGFKLIDVEIIFNERSGSGRPFPENSADDRAREKHTG
VEGRKLEVNNAATVMTNKTVPNTPTNGKLNIPYVGAIVYSPDFAGVYLQANOG
SSMYSGPSLIVYTSAMGFPYPAATAAAYGAHLRGKRTVYTFRAAAPPPPIPAY
GVVYVDGFYADIVGGYAAAYRAQPTPATAAASDSYGRVYADPFRHTLAAPTGY
VGANNAFAPFLDADKTRSHADVDGVLSSLQASYIRGYNRPAPY"
misc_feature 358..375
/gene="A2bp"
/note="Region: RNA binding motif RNP-2"
misc_feature 475..495
/gene="A2bp"
/note="Region: RNA binding motif RNP-1"
BASE COUNT 362 a 396 c 324 g 281 t
ORIGIN
Alignment Scores:
Pred. No.: 2.64e-77 Length: 1363
Score: 100.00 Matches: 100
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 67.57% Indels: 0
DB: 10 Gaps: 0
US-09-809-545a-2_copy_1_148 (1-148) x AF107204 (1-1363)
QY 7 ValAsnProTyrThrAsnGlyTyrPheLeuAsnProValAlaGlyAlaValTyrSerPro 26
DB 613 GTCAACCCCTACACCAATGCGTGAATTAATCAAGTTGTGGCGCGGTCTACAGCCCC 672
QY 27 AspPheTyrAlaGlyThrValLeuLeuGlyGlnAlaAsnGlnGlySerSerMetTyr 46
DB 673 GACTTCTATGACGACGAGTGTGTTGTGCCAGGCCAACCAAGAGGATCTTCATGTAC 732
QY 47 SerGlyProSerSerLeuValTyrThrSerAlaMetProGlyPheProTyrProAlaAla 66
DB 733 AGTGGCCCGACGTTCACTTGTATATCTTCTGCAATGCCCTGCTCCCATATCCGGCGCC 792
QY 67 ThrAlaAlaAlaAlaTyrArgGlyValAlaHisLeuArgGlyArgThrValTyrAsn 86
DB 793 ACTGCTGCGAGCTGACATACCGAGGGCTCACCTTCAGGCGCGGTGCGACCGTGTACAA 852
QY 87 ThrPheArgAlaAlaAlaProProProProAlaTyrGlyGlyValAlaTyrGln 106
DB 853 ACCTTCAGGCTGACAGCCCCCGCCCAATCCCGGCTTATGCGGTGTGTTTACAG 912

```
RESULT 3
AF094849      1547 bp  mRNA      linear  PRI 01-MAR-2002
LOCUS
DEFINITION   Homo sapiens hexaribonucleotide binding protein 1 isoform alpha
              (HRNP1) mRNA, complete cds.
ACCESSION    AF094849
VERSION      AF094849.1  GI:19032365
KEYWORDS
SOURCE       Homo sapiens.
ORGANISM     Homo sapiens.
REFERENCE    1 (bases 1 to 1547)
AUTHORS      Chen, W., Chu, Z.-L., Blough, R. I., Liu, L., Hoppes, B. and
              Winkelmann, J. C.
TITLE        Molecular cloning and chromosomal localization of a human brain,
              heart and skeletal muscle specific RNA binding protein gene
              homologous to fox-1 in Caenorhabditis elegans
              Unpublished
              2 (bases 1 to 1547)
JOURNAL      Submitted (24-SEP-1998) Internal Medicine/Hematology-Oncology,
REFERENCE    University of Cincinnati College of Medicine, 231 Bethesda Ave.,
AUTHORS      Cincinnati, OH 45267-0508, USA
TITLE        Location/Qualifiers
JOURNAL
FEATURES
source
1. 1547
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="16"
   /map="16p13.3"
   1. 1547
   /gene="HRNP1"
   254..1432
   /gene="HRNP1"
   /note="similar to Caenorhabditis elegans fox-1"
   /codon_start=1
   /product="hexaribonucleotide binding protein 1 isoform
   alpha"
   /protein_id="AA183405.1"
   /db_xref="GI:19032366"
   /translation="MLASQGVLLHPYGVPMIVPAAPYLPGLIQNOEAAAADPTMAOP
   VASQAPAPONGIPAEYTAHPHAPRYGTQTVPEHTLNLVPAQTHSESPADTSA
   QVSGTATQTDAAPTDGOPTQPSSENTENKSKPKLHVSNIPIRFRDPDLROMGQF
   KTLIDVEITINERSGSKGFGVTENSDADBARAKLHGTVEGKIEVNNATAVMTN
   KKTVPNTNGKLNIPVVGAVYSPEFYAGTVLLCOANQSGSMYSAPSLVYTSAMPF
   PYPAATAAARGAHLRGRTVYNTFRAAAPPIIPAYGGVVYQEPVYGNKLLQGGY
   AAYRYQPTPATAAYSDSYGRVYAADPYHHALAPATYGVGMASIVRGYNRFAY
   "
BASE COUNT  380 a   448 c   374 g   345 t
ORIGIN
Alignment Scores:
Pred. No.:      2,91e-77      Length:      1547
Score:          100.00      Matches:      100
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    67.57%      Indels:      0
Gaps:          0
DB:
US-09-809-545A-2_COPY_1_148 (1-148) x AF094849 (1-1547)
QY  49  ProSeSerLeuValTyrThSerAlaMeCPProGlyPheProTyrProAlaAlaThrAla 68
Db  1043  CCCAGTTCACTTGATATATACCTTCGCAATGCCAGGCTTCCCTATCCAGCACCCGCC 1102
QY  69  AlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArgGlyArgThValTyrAsnThrPhe 88
Db  1103  GCGGCGGCTACCGAGGGCGGACCTCGGAGGCCGGGGGCGGACCGGTGACCAACCTTC 1162
QY  89  ATGAlaAlaAlaProProProProIleProAlaTyrGlyGlyValValTyrGlnGluPro 108
```

```
Db  1163  AGGCGCGGCGGCCCGCCCGCCCGATCCCGGCTACGGCGGAGTAGTATCAAGAGCCT 1222
QY  109  ValTyrGlyAsnLysLeuLeuGlnGlyGlyTyrAlaAlaTyrArgTyrAlaGlnProThr 128
Db  1223  GTGATGCGCAATTAATGTCTGCAGGCTGTTATGCTGATGCCCTACGCCACCGCTAC 1282
QY  129  ProAlaThrAlaAlaAlaTyrSerAspSerTyrGlyArgValTyrAlaAlaAspProTyr 148
Db  1283  CCTGCCACTGCGCGCTGCTACAGTACGACGAGAGAGTTATGCTGCGCACCCCTAC 1342
RESULT 4
AF229057      2000 bp  mRNA      linear  PRI 01-FEB-2002
LOCUS
DEFINITION   Homo sapiens hexaribonucleotide binding protein 1 isoform gamma
              (HRNP1) mRNA, complete cds.
ACCESSION    AF229057
VERSION      AF229057.1  GI:18461366
KEYWORDS
SOURCE       Homo sapiens.
ORGANISM     Homo sapiens.
REFERENCE    1 (bases 1 to 2000)
AUTHORS      Chen, W. and Winkelmann, J. C.
TITLE        Direct Submission
JOURNAL      Submitted (28-JAN-2000) Internal Medicine/Hematology-Oncology,
REFERENCE    University of Cincinnati College of Medicine, The Vontz Center for
AUTHORS      Molecular Studies, 3125 Eden Avenue, Cincinnati, OH 45267-0508, USA
TITLE        Location/Qualifiers
JOURNAL
FEATURES
source
1. 2000
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="16"
   /map="16p13.3"
   1. 2000
   /gene="HRNP1"
   256..1512
   /gene="HRNP1"
   /note="RNA-binding protein"
   /codon_start=1
   /product="hexaribonucleotide binding protein 1 isoform
   gamma"
   /protein_id="AA171904.1"
   /db_xref="GI:18461367"
   /translation="MLASQGVLLHPYGVPMIVPAAPYLPGLIQNOEAAAADPTMAOP
   VASQAPAPONGIPAEYTAHPHAPRYGTQTVPEHTLNLVPAQTHSESPADTSA
   QVSGTATQTDAAPTDGOPTQPSSENTENKSKPKLHVSNIPIRFRDPDLROMGQF
   KTLIDVEITINERSGKGFVTEENSDADBARAKLHGTVEGKIEVNNATAVMTN
   KKTVPNTNGKLNIPVVGAVYSPEFYAGTVLLCOANQSGSMYSAPSLVYTSAMPF
   PYPAATAAARGAHLRGRTVYNTFRAAAPPIIPAYGGVVYQEPVYGNKLLQGGY
   AAYRYQPTPATAAYSDSYGRVYAADPYHHALAPATYGVGMNAPAPLTDKTRSH
   ADVGVLSSLQASIVRGYNRFAPY"
BASE COUNT  502 a   505 c   487 g   506 t
ORIGIN
Alignment Scores:
Pred. No.:      3,55e-77      Length:      2000
Score:          100.00      Matches:      100
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    67.57%      Indels:      0
Gaps:          0
DB:
US-09-809-545A-2_COPY_1_148 (1-148) x AF229057 (1-2000)
QY  49  ProSeSerLeuValTyrThSerAlaMeCPProGlyPheProTyrProAlaAlaThrAla 68
Db  1045  CCCAGTTCACTTGATATATACCTTCGCAATGCCAGGCTTCCGATACCGACGCCGCC 1104
QY  69  AlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArgGlyArgThValTyrAsnThrPhe 88
Db  1105  GCGGCGGCTACCGAGGGCGGACCTCGGAGGCCGGGGTGCACCGGTGACCAACCTTC 1164
```

QY	89	ARGALAAALAAAPROPEPROPEPROLEAFROALATYRGYGLVAlValTYTGlnGlnUPRO	108
DB	1165	AGGGCCGGGGCCCCCCCCCGGATCCGGGCTACGGCGAGTGTATACAGAGCT	1224
QY	109	VAlTYRGYAsnLysLeuLeuGlnGlyTYTyrAlaAlaTYTArgTYrAlaGlnProThr	128
DB	1225	GGTATGGCAATAAATGTGTGCAGAGGTGTATGTGCTACACCGCTACGCCAGCTAC	1284
QY	129	PROAlaThrAlaAlaAlaTYTserKapsERTYRGYVArgValTYrAlaAlaAspProTYr	148
DB	1285	CCTGCACATCGCGCTGCCTACAGTAGCAGATTACGACGATTATGTGCGACCCCTAC	1344
RESULT 5			
AB041596			
LOCUS	AB041596	2002 bp	mRNA linear ROD 30-JUN-2000
DEFINITION	Mus musculus brain cDNA, clone MNCD-3035, similar to Homo sapiens		
ACCESSION	AB041596		
VERSION	AB041596.1 GI:7670455		
KEYWORDS	file (full insert sequence)		
SOURCE	Mus musculus (strain:C57BL/6) adult female cDNA to mRNA, clone_11b:Sugano mouse brain mncb clone:MNCB-3035.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (sites) Osada,N., Kuesuda,J., Tanuma,R., Ito,A., Hirata,M., Sugano,S. and		
JOURNAL	Hashimoto,K.		
REFERENCE	Isolation of full-length cDNA clones from mouse brain cDNA library		
AUTHORS	made by oligo-capping method		
TITLE	Unpublished		
JOURNAL	2 (bases 1 to 2002)		
REFERENCE	Hashimoto,K., Osada,N., Kuesuda,J. and Sugano,S.		
AUTHORS	Direct Submission		
TITLE	Submitted (12-APR-2000) Katsuyuki Hashimoto, National Institute of		
JOURNAL	Infectious Diseases, Division of Genetic Resources; 23-1, Toyama		
REFERENCE	1-chome, Shijuku-ku, Tokyo 162-8640, Japan		
AUTHORS	(E-mail:khashim@nih.go.jp, URL:http://www.nih.go.jp/yoken/genebank/		
TITLE	Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)		
JOURNAL	URL: http://www.nih.go.jp/yoken/genebank/		
REFERENCE	Lib Name: Sugano mouse brain mncb		
AUTHORS	Lab host: TOP10		
TITLE	Vector: pME18S-FL3		
JOURNAL	1st strand cDNA was primed with an oligo(dT) primer		
REFERENCE	[ATGTGGCTTTTATTTTTTTTTT]; double-stranded cDNA was synthesized		
AUTHORS	using specific 5' and 3' primers and amplified by PCR. The PCR		
TITLE	product was digested with SfiI and size selection was performed to		
JOURNAL	exclude fragments <1.5kb. The SfiI-digested PCR product was cloned		
REFERENCE	into distinct DraIII sites of pME18S-FL3. XhoI sites just outside		
AUTHORS	the DraIII sites can be used to isolate the cDNA insert. Library		
TITLE	was constructed by Sugano et al. (University of Tokyo, Institute of		
JOURNAL	Medical Science). Custom primer used for sequencing (5' end primer		
REFERENCE	[CTTCTGCTCTAAAGCTGGG]; 3' end primer		
AUTHORS	[CAGCTGACGCTGACGACAC]).		
TITLE	A part of this sequence is reported in AU067167.		
JOURNAL	Location/Qualifiers		
REFERENCE	1..2002		
AUTHORS	/organism="Mus musculus"		
TITLE	/strain="C57BL/6"		
JOURNAL	/db_xref="taxon:10090"		
REFERENCE	/clone="MNCB-3035"		
AUTHORS	/sex="female"		
TITLE	/clone_11b="Sugano mouse brain mncb"		
JOURNAL	/dev stage="adult"		
REFERENCE	637..1827		
AUTHORS	/note="unnamed protein product"		
TITLE	/codon_start=1		
JOURNAL	/protein_id="BAA95079.1"		
REFERENCE	/db_xref="GI:7670456"		
AUTHORS	/translat="MNCEROLRGNOEAAAPDTMAOPYASGAPFONGIPATYATP		
TITLE	HPHAPAYTGTTGTTGHTLNTLPPTGOSBOASPTASOTVSGTATDNDADAPTGGGO		
JOURNAL	TOPSENTSKSPKRLHVSINPFRPRDLDLQMGQPKLIDVETITNENSGKFGV		
REFERENCE	TFENADDDRAEKRLHGTVEGRKLEVNNAARVMTNKTNPYTNKRLNLPVGAAY		

BASE COUNT	518 a	589 c	516 g	379 t
ORIGIN	FAPY"			
Alignment Scores:				
Prod. No.:	3,55e-77	Length:	2002	
Score:	100.00	Matches:	100	
Percent Similarity:	100.00%	Conservative:	0	
Best Local Similarity:	100.00%	Mismatch:	0	
Query Match:	67.57%	Indels:	0	
DB:	10	Gaps:	0	
US-09-809-545A-2_COPY_1_148 (1-148) x AB041596 (1-2002)				
Qy	7	ValaenProTyTrThAsnGlyTrpLysLeuAsnProValValGlyAlaValTySerPro	26	
Db	1237	GTCAACCCCTACACCAATAGCGTGAATAATTAATCAGTTGTGGCGCGGTCTACAGCCCC	1296	
Qy	27	AappHeTyTrAlaGlyThValleuLeuCyseGlnAlaAngInGlySerSerMetTyTr	46	
Db	1297	GACTTCATATGACGAGCAGCGTGTCTGTGTGCCAGGCAACAGAGAGGATCTTCCATGTAC	1356	
Qy	47	SerGlyProSerSerLeuValTyTrHisSerAlaMetProGlyPheProTyProAlaAla	66	
Db	1357	AGTGCGCCCGATTCCTGTATATCTCTGCAATGCTTGTCCATATCCGCGCCCC	1416	
Qy	67	ThrAlaAlaAlaAlaTyTrArgGlyValHisLeuArgGlyArgGlyValTyTrAsn	86	
Db	1417	ACTGCTGACGCTGCATACGAGGGGCTACCTTCGAGGCCGCTGTCGACCGTATCAAC	1476	
Qy	87	ThrPheArgAlaAlaAlaProProProProIleProAlaTyTrGlyGlyValValTyTrGln	106	
Db	1477	ACCTTCAGGCGTGCAGCGCCCCCGCCCAATCCCGGCTATGCGGCGTGTATTACGAG	1536	
RESULT 6				
AF109106	AF109106	3348 bp	mRNA	linear PRI 01-MAR-2002
LOCUS	Homo sapiens hexaribonucleotide binding protein 1 isoform beta			
DEFINITION	(HNRNP1) mRNA, complete cds.			
ACCESSION	AF109106			
VERSION	AF109106.1 GI:19032367			
KEYWORDS				
SOURCE	Homo sapiens.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	1 (bases 1 to 3348) Chen,W., Chu,Z.-L., Blough,R.I., Liu,L., Hoppes,B. and Winkelmann,J.C.			
TITLE	Molecular Cloning and Chromosomal Localization of a Novel Human Brain, Heart and Skeletal Muscle Specific RNA Binding Protein Gene Homologous to fox-1 in Caenorhabditis elegans			
JOURNAL	Unpublished			
AUTHORS	2 (bases 1 to 3348) Chen,W., Chu,Z.-L., Blough,R.I., Liu,L., Hoppes,B. and Winkelmann,J.C.			
TITLE	Direct Submission			
JOURNAL	Submitted (24-NOV-1998) Internal Medicine/Hematology-Oncology, ML University of Cincinnati College of Medicine, 231 Bethesda Ave. ML 0508, Cincinnati, OH 45267-0508, USA			
FEATURES	Location/Qualifiers			
source	1..3348			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/chromosome="16"			
	/map="16p13.3"			
gene	1..3348			
	/gene="HNRNP1"			
	254..1441			
CDS	/gene="HNRNP1"			

```
/note="similar to Caenorhabditis elegans fox-1"
/codon_start=1
/product="hexaribonucleotide binding protein 1 isoform
beta"
/protein_id="AA183406.1"
/db_xref="GI:19032368"
/translaction="MLASQGVLLHPYGVPMIVPAAPYLPGLQGNQEAAPPTMAOP
VSAQAPAPONGI PAEYTAHPHAPPEYTGQTVPHPTLNLYPQTHSEQSPADTSA
QVSGATQTDAAPTDGPQOTQPOSEVENTESKOPKRLHVSNI PRFRDPLRQMFQCF
GKILDEVIIFENERGSKGFVTPENSADARAREKLGTVVEGRKIEVNNAATAMVN
KTVNPTNCKMKNLPVGAAYSPFVAGTULLCOANQSGSSMVSAPSILVYTSAMGF
PYPAATRAAAIRGAHLRGRTGYNTFRRAAPPPIPAVGGVYQEPVYGNKLQGGTA
AAYRAQPTPATAAAYSDRQFVFAADEISCNISAVTDEFVLPTTTHLLQPPPTA
LVP"
polysignal
3299..3304
/gene="HRBp1"
BASE COUNT 950 a 775 c 711 g 912 t
ORIGIN
Alignment Scores:
Pred. No.: 1,71e-67 Length: 3348
Score: 89.00 Matches: 89
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 60.14% Indels: 0
DB: 9 Gaps: 0
US-09-809-545a-2_COPY_1_148 (1-148) x AF109106 (1-3348)
QY 49 ProSerSerLeuValIYrThSerAlaMetProGlyPheProTYrProAlaAlaThrAla 68
Db 1043 CCCAGTTCACCTGTATATCTTCTGCAATGCGCAGGCTTCCGTAATCCAGCAGCCGCC
QY 69 AlAlaAlaIYrArgGlyAlaHisLeuArgGlyArgGlyArgGlyArgGlyArgGly 88
Db 1103 GCGGCCCCCTACCCAGGGGCGCCAGGCGCGGCTCCAGCGGTACAAACACCTTC
QY 89 ArgAlaAlaIaProProProProIleProAlaTYrGlyGlyAlaIYrGlnIuPro 108
Db 1163 AGGCGCGGCGGCGGCGGCGGCGGCGGCTCCAGCGGAGTAGTATACAGAGGCT 1222
QY 109 ValTYrGlyAsnLYLeuLeuGlnGlyTYrAlaAlaIYrArgTYrAlaGlnProThr 128
Db 1223 GTGATGGAATTAATGCTCAGGGTGTATGCTGATACCGCTACGCCACGCTACC 1282
QY 129 ProAlaThrAlaAlaIYrSerAsp 137
Db 1283 CCTGCCACCTGCGCTGCTACAGTGAC 1309
RESULT 7
AF191501 1586 bp mRNA linear ROD 01-MAR-2002
LOCUS AF191501
DEFINITION Mus musculus hexaribonucleotide binding protein 1 (Hrbp1) mRNA,
complete cds.
VERSION AF191501
KEYWORDS AF191501.1 GI:19032413
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1586)
Chen, W. and Winkelman, J. C.
Direct Submission
Submitted (01-OCT-1999) Internal Medicine/Hematology-Oncology,
University of Cincinnati College of Medicine, 231 Bethesda Ave.,
Cincinnati, OH 45267-0508, USA
FEATURES
source
location/Qualifiers
1..1586
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="skeletal muscle"
1..1586
gene
```

```
/gene="Hrbp1"
CDS
166..1419
/gene="Hrbp1"
/codon_start=1
/product="hexaribonucleotide binding protein 1"
/protein_id="AA183425.1"
/db_xref="GI:19032414"
/translaction="MLASQGVLLHSHYGVPMIVPAAPYLPGLQGNQEAAPPTMAOP
VSAQAPAPONGI PAEYTAHPHAPPEYTGQTVPHPTLNLYPQTHSEQSPADTSAO
QVSGATQTDAAPTDGPQOTQPOSEVENTESKOPKRLHVSNI PRFRDPLRQMFQCF
GKILDEVIIFENERGSKGFVTPENSADARAREKLGTVVEGRKIEVNNAATAMVN
KTVNPTNCKMKNLPVGAAYSPFVAGTULLCOANQSGSSMVSAPSILVYTSAMGF
PYPAATRAAAIRGAHLRGRTGYNTFRRAAPPPIPAVGGVYQEPVYGNKLQGGTA
AAYRAQPTPATAAAYSDSYGRVYAADPYHHTTLAPATYVGAAATAPATDARTSHA
DVGVLVLSLQASLYRGYNRPAPY"
BASE COUNT 407 a 458 c 374 g 347 t
ORIGIN
Alignment Scores:
Pred. No.: 1.13e-42 Length: 1586
Score: 60.00 Matches: 139
Percent Similarity: 95.86% Conservative: 0
Best Local Similarity: 95.86% Mismatches: 3
Query Match: 40.54% Indels: 6
DB: 10 Gaps: 0
US-09-809-545a-2_COPY_1_148 (1-148) x AF191501 (1-1586)
QY 7 ValAsnProTYrThraNGIYrTbLYLeuAsnProValGlyAlaValTYrSerPro 26
Db 826 GTCAACCCCTACACCAATGCTGGAATTAATCAATGTGGCGCGGTACAGCCCC
QY 27 AspPheTYrAlaGlyThraValLeuLeuCYGlnAlaAsnGlnGlySerSerMetTYr 46
Db 886 GACTTCTATGAGGCAAGGCTGCTTGTGCCAGGCCAACAGAGGATCTTCATGTAC 945
QY 47 SerGlyProSerSerLeuValIYrThSerAlaMetProGlyPheProTYrProAlaAla 66
Db 946 AGTGCCCGCATCTGTATATCTTCTGCAATGCTGCGCTTCCATATCCGCGGCC
QY 67 Thr-AlaAlaAlaIYrArgGlyAlaHisLeuArgGlyArgGlyArgGlyArgGly 86
Db 1006 TC-TGCTGACGCTGCACACGAGGGGCTCACCTTCGAGCGGTGCGACCGG-GTACA 1063
QY 86 snTrPhe-ArgAlaAlaIaProProProProIleProAlaTYrGlyGlyAlaIYr 105
Db 1064 ACACCTT-CAGGCTGCAAGCGGCGGCGGCGGCGGCTTATGGCGAGTAGTAT 1122
QY 106 GlnIuProValTYrGlyAsnLYLeuLeuGlnGlyTYrAlaAlaIYrArgTYrAla 125
Db 1123 CAAGAGCAGATGATGCAATTAATGCTCAGGGTGTATGCTGCGTACCGCTATGCC 1182
QY 126 GlnProThrProAlaThrAlaAlaIYrSerAspSerTYrGlyArgValTYrAlaAla 145
Db 1183 CAGCCCAACCCCTGCGCTGCGCTACAGTAGTACAGGACGATTATGCTGCC 1242
QY 146 AspProTYr 148
Db 1243 GACCCCTAC 1251
RESULT 8
AB060859 1475 bp mRNA linear PRI 13-JUN-2001
LOCUS AB060859
DEFINITION Macaca fascicularis brain cDNA clone: Qcra-11594, full insert
sequence.
ACCESSION AB060859.1 GI:13874510
VERSION AB060859.1
KEYWORDS oligo capping; fie (full insert sequence).
SOURCE Macaca fascicularis adult male temporal lobe right CDNA,
clone 11b:macaque brain cDNA library Qcra clone:Qcra-11594.
Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
```

REFERENCE	Cercopithecinae; Macaca.
AUTHORS	Oseada, N., Hida, M., Kuseida, J., Tanuma, R., Ieki, K., Hirai, M., Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K.
TITLE	Isolation of full-length cDNA clones from macaque brain cDNA libraries
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 1475)
AUTHORS	Hashimoto, K., Oseada, N., Hida, M., Kuseida, J. and Sugano, S.
TITLE	Direct Submission
JOURNAL	Submitted (27-Apr-2001) Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
REFERENCE	(E-mail: khashi@nih.go.jp, URL: http://www.nih.go.jp/yoken/genebank/, Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)
COMMENT	Lab host: TOPI0 Vector: pME18S-FL3 (Acc. No. AB009864) R. Site1: DraIII (CACTGTGTG) R. Site2: DraIII (CACCATGTG) Description: 1st strand cDNA was primed with an oligo (dt) primer (ATGTGGCTTTTCTTTTCTTTT); double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18S-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by oligo-capping method (Sugano et al., , Institute of Medical Science, University of Tokyo).
FEATURES	Custom primer used for sequencing (5' end primer [CTTCTGCTCTAAAGTCGCG] ; 3' end primer [CGACCTGCAGCTCGACGACAC]). Location/Qualifiers 1..1475 /organism="Macaca fascicularis" /db_xref="taxon:9541" /clone="QCR-11594" /sex="male" /issue_type="temporal lobe right" /clone_lib="macaque brain cDNA library QCR" /dev_stage="adult" 90..1220 /codon_start=1 /product="hypothetical protein" /protein_id="BAB46877.1" /db_xref="GI:13874511" /translation="MEEKSGSRMVOQNOEAAAADPTMAQPYASQAPAPONGIPAEYVY APHPAPAEYGGQTVTPHRLNLVPAQPAETHESPADNAQTGATQTDADAPDGG QPOQPSSENTENDKSOQPKLHYSNI PFRPRDKDLQMGQPGKILIDVELI ENERSKGG GFTVFPENSADADRAEKRLHGTVNERSKRL EVNNATRAVRWNTKNTYPTNGKLNAPVYG AVSPEFPAAGTVLDCQANQESSMSASVLSVTSAMGFPPYPATATAAYRGALHGG RGRVTVNFRFAAAPPPPIPAYGGVYQDGFAGDIYGGVAYVRAQPPATVAAVSDR NQFVVADEISCTNSAVTDEFMFLPTTTHLLOPPPALVP"
CDS	389 a 414 c 363 g 309 t
ORIGIN	
Alignment Scores:	
Pred. No.:	5.74e-41 Length: 1475
Score:	58.00 Matches: 58
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	39.15% Indels: 0
Gaps:	0
US-09-809-545A-2_COPY_1_148 (1-148) x AB060859 (1-1475)	
QY	49 ProSerSerLeuValTyThrSerAlaMetProGlyPheProTyPProAlaAlaThrAla 68
DB	825 CCCAGTCACTGTGATATACCTCTGCAGAACGACGCTCCGATACACACACCGCC 884
QY	69 AAlaAlaATyArGgLyAlaHsleuAATgGcYArGgLYArGThValTyRanThrPhe 88
DB	885 GGCGCGCGCTACCGAGGGCGCACCTGCAGAGCGCGGGCGGACGCTGTACAACACTTC 944

```

Oy      89  ArgAlaAlaAlaProProProProlerolAlaTyrGlyGlyValValTyrGln 106
Db      945  AGGGCCGAGGCCCCCCCCCATCCGGCCCTATGCGGCGTGTGTTTACAG 998

RESULT 9
AK001027      1513 bp      mRNA      linear      PRI 01-AUG-2002
AK001027      LOCUS      Homo sapiens cDNA FLJ10165 fis, clone HEMBA1003591, weakly similar
DEFINITION
ACCESSION      AK001027
VERSION        AK001027.1 GI:7022045
KEYWORDS       oligo capping; fis (full insert sequence).
SOURCE         Homo sapiens embryonic, 10 weeks whole embryo, mainly head cDNA to
               mRNA, clone lib:HEMBAL clone:HEMBAL003591.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS        Isegai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
               Nishikawa,T., Nagai,K., Sugano,S., Shitatori,A., Sudo,H.,
               Magatsuma,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H.,
               Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
               Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
               Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
               Nakamura,Y., Nagaharti,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
               NEDO human cDNA sequencing project
TITLE          Unpublished
JOURNAL        2 (bases 1 to 1513)
REFERENCE      Isegai,T. and Otsuki,T.
AUTHORS        Direct Submission
TITLE          Submitted (16-FEB-2000) Takao Isegai, Helix Research Institute,
               Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
               (E-mail:genomics@hri.co.jp, Tel:81-438-52-1975, Fax:81-438-52-1986)
JOURNAL        NEDO human cDNA sequencing project supported by Ministry of
               International Trade and Industry of Japan. cDNA full insert
               sequencing: Research Association for Biotechnology, cDNA library
               construction, 5'- & 3'-end one pass sequencing and clone selection:
               Helix Research Institute (supported by Japan Key Technology Center
               etc.) and Department of Virology, Institute of Medical Science,
               University of Tokyo.

FEATURES
             source
             1..1513
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="HEMBAL003591"
                /tissue_type="whole embryo, mainly head"
                /clone_lib="HEMBAL"
                /dev_stage="embryo, 10 weeks"
                /note="cloning vector: pMB18SFL13"
             221..1414
                /note="unamed protein product"
             /codon_start=1
             /protein_id="BAA91472.1"
             /db_xref="GI:7022046"
             /translation="MNCERBQLRGNQAAAAAPDTMAQPYASQFAPRPNQCIPIAETAR
             HPHPAPEYTGQTVPEHTLNLTPPAQTHSEQSPADTSAQVTSGRATQADADAPTDGQ
             QTQSENTGNSQPKRLHVSNIPEFRPDPDRLQWFGFKLIDVEILFENSGSGFGGA
             VTFPNSADADARAEKLGTVVEGRKLEVNNAATSMKNTVNTGMLKNPVGAAGAA
             VSPFVAGTVILCOANQSGSMVSPASPLVYTSAMPGPYPAAATAAAYRAHLRGRAV
             RTVNTPRRAAPRPPIPAVGCVVVQDFGADIDGVAAYVADQPTATAAYSDSYG
             RVYADPHTHAALPAPTYGAGNMAFPPLDAKTRISADVDGVLVLSLQASLYNGGVYV"
             RFAV"

BASE COUNT      354 a      418 c      395 g      346 t
ORIGIN
Alignment Scores:
Pred. No.:      5,86e-41      Length:      1513
Score:          58.00         Matches:      58
Percent Similarity: 100.00%   Conservative: 0
Base Local Similarity: 100.00% Mismatches:      0
Query Match:     39.19%      Indels:         0
DB:              Gaps:         0

```


US-09-809-545a-2_COPY_1_148 (1-148) x AK001027 (1-1513)

Qy 49 ProSerSerLeuValIYrThrSerAlaMetProGlyPheProIYrProAlaAlaThrAla 68
Db 950 CCCAGTTCACCTTGATATACCTTCTGCAATGCCAGGCTCCGATCCAGCACCCGCC 1009

Qy 69 AAlaAlaIaTyArgGlyAlaHisLeuArgGlyArgGlyArgThrValIYrAsnThrPhe 88
Db 1010 GCGGCGCGCTACCGAGGGGCGACCTCGAGGCGCGGTCGACCGTGTACACACCTTC 1069

Qy 89 ArgAlaAlaIaIaIaProProProProIleProAlaIaTyGlyGlyValIaIYrGln 106
Db 1070 AGGCGCGGCGCGCCGCCGCCGATCCGCGCTACGCGCGGTGTATTACAG 1123

RESULT 10
AF107203

LOCUS AF107203 2279 bp mRNA linear PRI 23-JUN-2000
DEFINITION Homo sapiens ataxin 2-binding protein (A2BP) mRNA, complete cds.
ACCESSION AF107203
VERSION AF107203.1 GI:8671585
KEYWORDS

SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 2279)
Shibata, H., Huynh, D. P., Vo, T. T. and Pulst, S.-M.
A novel protein, A2BP, with RNA binding motif binds to C-terminal
ataxin-2
Unpublished
2 (bases 1 to 2279)
Shibata, H., Huynh, D. P., Vo, T. T. and Pulst, S.-M.
Direct Submission
Submitted (16-NOV-1998) Division of Neurology and Neurogenetics
Laboratory, Burns and Allen Research Institute, Cedars-Sinai
Medical Center, UCLA School of Medicine, 8700 Beverly Blvd., Los
Angeles, CA 90048, USA
Location/Qualifiers
1..2279
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="similar to Homo sapiens cDNA clone HSP38A20"
1..2279
/gene="A2BP"
1047..2180
/gene="A2BP"
/codon_start=1
/product="ataxin 2-binding protein"
/protein_id="AAF78291.1"
/db_xref="GI:8671586"
/translation="MAQPIYASQAPAPONGIPATYTPAHPHAPETTGTTTPEHTLN
LYPRAPQTHSDPADTSAGTVSGATOTDDAPTDGDPOTOPSTNTKSOQKRLHVS
NIPRFRPDRLQMFQGFGLIDVEIIENRSGSGFGFVFEENSDRAREKLAGYV
VEGKRIEIVNNAATVNTKNTVNTKNTVNTKNTVNTKNTVNTKNTVNTKNTVNTKNTV
SMYGAPSLSLYTSAMPGRPPYPAATAAAYKYNLPGVAVSPEFYAGTVLLCOAQEES
GVVQDGFYGDADIVGYAAVRYADPTATTAAYSDVGRVYADPHYHALAPATYGV
GAMNAFAPLIDAKTRSHADVGLVLSLQASIVRGYNNRPAPY"

misc_feature
1341..1358
/gene="A2BP"
/note="Region: RNA binding motif RNP-2"
1452..1472
/gene="A2BP"
/note="Region: RNA binding motif RNP-1"

BASE COUNT 497 a 712 c 636 g 434 t
ORIGIN

Alignment Scores:
Pred. No.: 8.03e-41 Length: 2279
Score: 58.00 Matches: 58
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 39.19% Indels: 0

DB: 9 Gaps: 0

US-09-809-545a-2_COPY_1_148 (1-148) x AF107203 (1-2279)

Qy 49 ProSerSerLeuValIYrThrSerAlaMetProGlyPheProIYrProAlaAlaThrAla 68
Db 1716 CCCAGTTCACCTTGATATACCTTCTGCAATGCCAGGCTCCGATCCAGCACCCGCC 1775

Qy 69 AAlaAlaIaTyArgGlyAlaHisLeuArgGlyArgGlyArgThrValIYrAsnThrPhe 88
Db 1776 GCGGCGCGCTACCGAGGGGCGACCTCGAGGCGCGGTCGACCGTGTACACACCTTC 1835

Qy 89 ArgAlaAlaIaIaIaProProProProIleProAlaIaTyGlyGlyValIaIYrGln 106
Db 1836 AGGCGCGGCGCGCCGCCGCCGATCCGCGCTACGCGCGGTGTATTACAG 1889

RESULT 11
AR134676

LOCUS AR134676 2372 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 1 from patent US 6194171.
ACCESSION AR134676
VERSION AR134676.1 GI:14123581
KEYWORDS

SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
1 (bases 1 to 2372)
Pulst, S.M. and Shibata, H.
Nucleic acids encoding ataxin-2 binding proteins
Patent: US 6194171-A 1 27-FEB-2001;
Location/Qualifiers
1..2372
/organism="Unknown"

BASE COUNT 548 a 726 c 646 g 452 t
ORIGIN

Alignment Scores:
Pred. No.: 6.06e-40 Length: 2372
Score: 57.00 Matches: 57
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 38.51% Indels: 0
DB: Gaps: 0

US-09-809-545a-2_COPY_1_148 (1-148) x AR134676 (1-2372)

Qy 49 ProSerSerLeuValIYrThrSerAlaMetProGlyPheProIYrProAlaAlaThrAla 68
Db 1716 CCCAGTTCACCTTGATATACCTTCTGCAATGCCAGGCTCCGATCCAGCACCCGCC 1775

Qy 69 AAlaAlaIaTyArgGlyAlaHisLeuArgGlyArgGlyArgThrValIYrAsnThrPhe 88
Db 1776 GCGGCGCGCTACCGAGGGGCGACCTCGAGGCGCGGTCGACCGTGTACACACCTTC 1835

Qy 89 ArgAlaAlaIaIaIaProProProProIleProAlaIaTyGlyGlyValIaIYrGln 105
Db 1836 AGGCGCGGCGCGCCGCCGCCGATCCGCGCTACGCGCGGTGTATTAC 1886

RESULT 12
HSFOX09

LOCUS HSFOX09 251 bp DNA linear PRI 01-MAR-2002
DEFINITION Homo sapiens hexaribonucleotide binding protein 1 (HRNB1) gene,
exon 9.
ACCESSION AF109115
VERSION AF109115.1 GI:19032378
KEYWORDS

SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 251)
Chen, W. and Winkelman, J. C.

JOURNAL	The Exon-Intron Organization of the Human HRNB1 Gene
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 251) Chen,W., Chu,Z.-L., Blough,R.I., Liu,L., Hoppe,B. and Winkelmann,J.C.
TITLE	Molecular Cloning and Chromosomal Localization of a Novel Human Brain, Heart and Skeletal Muscle Specific RNA Binding Protein Gene Homologous to fox-1 in Caenorhabditis elegans
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 251)
AUTHORS	Chen,W. and Winkelmann,J.C.
TITLE	Direct Submission Submitted (29-NOV-1998) Internal Medicine/Hematology-Oncology, University of Cincinnati College of Medicine, 631 Bethesda Ave., ML 0508, Cincinnati, OH 45267-0508, USA
FEATURES	Location/Qualifiers
source	1..251 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="16" /map="16p13.3" 67..199 /gene="HRNBPI" /number=9
exon	
BASE COUNT	29 a 111 c 70 g 41 t
ORIGIN	
Alignment Scores:	
Pred. No.:	1.86e-29 Length: 251
Score:	44.00 Matches: 44
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	29.73% Indels: 0
DB:	Gaps: 0
US-09-809-545A-2_COPY_1_148 (1-148) x HSFOX09 (1-251)	
Oy	59 ProGlyPheProTyrTrpAlaIalaThrAlaIalaIalaIalaTyArgLysAlaHisLeuArg 78
Dd	69 CCAGCCTTCCTCATCCACGACCACCGCGGGCGGCCTTACGAGGCGGCACTTGGA 128
Oy	79 GtATGAGtATgATThValTYrAnThPhEAATGALALAIAAIPROpROpRoiIlepro 98
Dd	129 GCCCGCGGTGGACCGGTGTACAACACTTCAGAGCGCGGCCGCCGCCGCCGATCCG 188
Oy	99 AAtATgATgATgLY 102
Dd	189 GCCTACGGCGGG 200
RESULT 13	
LOCUS	AC005774 164538 bp DNA linear PRI 05-OCT-1998
DEFINITION	AC005774/c Homo sapiens chromosome 16, BAC clone 2603 (HLNLI), complete sequence.
ACCESSION	AC005774
VERSION	AC005774.1 GI:3702264
KEYWORDS	HTG.
SOURCE	Homo sapiens. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
ORGANISM	
REFERENCE	LARGE SCALE SEQUENCE ANALYSIS AND ANNOTATION WITH THE SEQUENCE COMPARISON ANALYSIS (SCAN) SYSTEM
AUTHORS	Unpublished
TITLE	2 (bases 1 to 164538) Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L., Bryant,J., Tesmer,J., Melnick,L., Longmire,J., White,S., Ueng,S., Tatum,O., Campbell,C., Faggett,J., Matibale,M., and Deaven,L.
JOURNAL	Sequencing of Human Chromosome 16p13.3

JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 164538)
AUTHORS	Ricke,D.O., Bruce,D., Muntz,M., Doggett,N., Munk,C., Saunders,E., Robinson,D., Jones,M., Buckingham,J., Chasreen,L., Thompson,S., Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Yeng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M. and Deaven,L.
TITLE	Direct Submission
JOURNAL	Submitted (05-OCT-1998) Center for Human Genome Studies, DOE Joint Genome Institute, Los Alamos National Laboratory, MS M888, Los Alamos, NM 87545, USA
FEATURES	Location/Qualifiers
source	1. .164538
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="16"
	/map="16p13.3"
	/clone="2603"
repeat_region	1378. .1665
	/rpt_family="Alu"
repeat_region	complement(2144. .2320)
	/rpt_family="MER21"
repeat_region	complement(2198. .2416)
	/rpt_family="MER33"
repeat_region	complement(2450. .2725)
	/rpt_family="Alu"
repeat_region	complement(3157. .3393)
	/rpt_family="Alu"
repeat_region	complement(3524. .3792)
	/rpt_family="Alu"
repeat_region	5731. .6020
	/rpt_family="Alu"
repeat_region	complement(7557. .7727)
	/rpt_family="MER4"
repeat_region	8282. .8928
	/rpt_family="LTR8"
repeat_region	10055. .10310
	/rpt_family="THE1"
repeat_region	10472. .10618
	/rpt_family="L1"
repeat_region	10639. .10939
	/rpt_family="Alu"
repeat_region	complement(11109. .11357)
	/rpt_family="Alu"
repeat_region	12050. .12312
	/rpt_family="Alu"
repeat_region	complement(14642. .14755)
	/rpt_family="L1"
repeat_region	17429. .18370
	/rpt_family="MSTAR"
repeat_region	complement(19470. .19819)
	/rpt_family="MT1"
repeat_region	complement(20352. .20627)
	/rpt_family="Alu"
repeat_region	23808. .24058
	/rpt_family="Alu"
repeat_region	27654. .28217
	/rpt_family="L1"
repeat_region	complement(32764. .32956)
	/rpt_family="L1"
repeat_region	complement(33392. .33503)
	/rpt_family="Alu"
repeat_region	34586. .34870
	/rpt_family="Alu"
repeat_region	34936. .35489
	/rpt_family="MT2B2"
repeat_region	complement(36940. .37172)
	/rpt_family="Alu"
repeat_region	37173. .37437
	/rpt_family="THE1"
repeat_region	41428. .41745
	/rpt_family="THE1"
repeat_region	complement(42057. .43927)

```

repeat_region /rpt_family="r1"
43924. .44555
repeat_region /rpt_family="r1"
45128. .45367
repeat_region /rpt_family="Alu"
45467. .45704
repeat_region /rpt_family="Alu"
46192. .46578
repeat_region /rpt_family="MER25"
complement(46779. .47141)
repeat_region /rpt_family="THE1"
complement(48265. .48550)
repeat_region /rpt_family="Alu"
49630. .50221
repeat_region /rpt_family="MER41"
50781. .51096
repeat_region /rpt_family="Alu"
complement(51219. .51439)
repeat_region /rpt_family="Alu"
51462. .51708
repeat_region /rpt_family="MER33"
complement(52042. .52321)
repeat_region /rpt_family="Alu"
54001. .54092
repeat_region /rpt_family="Alu"
complement(55760. .56014)
repeat_region /rpt_family="Alu"
56616. .56899
repeat_region /rpt_family="Alu"
complement(66463. .66747)
repeat_region /rpt_family="Alu"
complement(66908. .67180)
repeat_region /rpt_family="Alu"
complement(70389. .70704)
repeat_region /rpt_family="MER1"
complement(72166. .72421)
repeat_region /rpt_family="MER33"
72501. .72769
repeat_region /rpt_family="Alu"
72643. .72956
repeat_region /rpt_family="MER5"
complement(76503. .76564)
repeat_region /rpt_family="Alu"
97534. .97700
repeat_region /rpt_family="MER3"
complement(98469. .98526)
repeat_region /rpt_family="Alu"
complement(99089. .99381)
repeat_region /rpt_family="Alu"
101648. .101930
repeat_region /rpt_family="Alu"
102790. .103056
repeat_region /rpt_family="Alu"
104010. .104277
repeat_region /rpt_family="Alu"
complement(111121. .111263)
repeat_region /rpt_family="MER5"
complement(112263. .112476)
repeat_region /rpt_family="MER30"
112956. .113234
repeat_region /rpt_family="Alu"
114385. .114657
repeat_region /rpt_family="Alu"
complement(117031. .117194)
repeat_region /rpt_family="MER20"
complement(118464. .118770)
repeat_region /rpt_family="Alu"
119271. .119549
repeat_region /rpt_family="Alu"
complement(122831. .122940)
repeat_region /rpt_family="MER41"
123471. .123742
repeat_region /rpt_family="Alu"

```

```

repeat_region complement(129860. .129972)
repeat_region /rpt_family="Alu"
135235. .135310
repeat_region /rpt_family="MLT1"
135874. .136157
repeat_region /rpt_family="Alu"
139775. .140048
repeat_region /rpt_family="Alu"
141734. .141814
repeat_region /rpt_family="MIR"
143071. .143352
repeat_region /rpt_family="Alu"
148286. .148773
repeat_region /rpt_family="MER1"
150431. .150707
repeat_region /rpt_family="Alu"
155151. .155387
repeat_region /rpt_family="Alu"
complement(156101. .156213)
repeat_region /rpt_family="MER20"
complement(156636. .156920)
repeat_region /rpt_family="Alu"
complement(159633. .159931)
repeat_region /rpt_family="Alu"
complement(160378. .160658)
repeat_region /rpt_family="Alu"
161617. .162140
repeat_region /rpt_family="Alu"
complement(163200. .163315)
repeat_region /rpt_family="MIR"
BASE COUNT 48114 a 34256 c 33738 g 48430 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 2.77e-27 Length: 164538
Score: 44.00 Matches: 44
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 29.73% Indels: 0
DB: 9 Gaps: 0

```

US-09-809-545A-2_COPY_1_148 (1-148) x ACC005774 (1-164538)

QY 59 ProGlyPheProTYrProAlaAlaThrAlaAlaAlaAlaTYrArgGlyAlaHisLeuArg 78

Db 118155 CAGGCTTCCCGTATCCAGCGACCGCCGCCGCTTACGAGGGGCCACCTGGCA 118096

QY 79 G1AyrG1yArGThrValTYrAsnThrPheArgAlaAlaAlaProProProPro11ePro 98

Db 118095 GGCGCGGTCCGACCGTGTACACACCTTCAGGCGCCGCCGCCGCCCGCATCCCG 118036

QY 99 A1ATyrG1yGly 102

Db 118035 GCCTACGGCGGG 118024

RESULT 14 HSA323688 596 bp DNA linear PRI 18-JUL-2002

LOCUS Homo sapiens genomic sequence surrounding NotI site, clone

DEFINITION N6-793C.

ACCESSION AJ323688

VERSION AJ323688.1 GI:15868067

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (bases 1 to 596)

AUTHORS

TITLE

Kutsenko, A.S., Gizatullin, R.Z., Al-Amin, A.N., Wang, F., Kvaasha, S.M.,
Podowski, R.M., Matshushin, Y.G., Gyanchandani, A., Muravenko, O.V.,
Levitsky, V.G., Kolchanov, N.A., Protopopov, A.I., Kashuba, V.I.,
Kiselev, L.L., Wasserman, W., Wahlstedt, C. and Zabarovsky, E.R.
NotI flanking sequences: a tool for gene discovery and verification

of the human genome
Nucleic Acids Res. 30 (14), 3163-3170 (2002)
12136039
2
2 (bases 1 to 596)
Zabarovskiy E.R.
Direct Submission
Submitted (16-MAY-2001) Microbiology and Tumorigenesis Centre,
Karolinska Institute, Theorells väg, 3, Box 266, Stockholm 171 77,
Sweden

BASE COUNT	132 a	171 c	158 g	133 t	2 others
------------	-------	-------	-------	-------	----------

US-09-809-545A-2_COPY_1_148 (1-148) X HSA323688 (1-596

Db 61 AGGGCGCGCGCGCGCGCGCGCGCGATCCCGGCTACGGCGGG 102

VERSION AJ335811.1 GI:15880229

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

of the human genome

Sweden

Source	1. .617
--------	---------

ORIGIN	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	

Pred. No.:	1.65e-20	Length:
------------	----------	---------

DB:	9	Gaps:	0
-----	---	-------	---

US-09-809-545A-2_COPY_1_148 (1-148) X HSA335811 (1-617)

Db 61 AGGGCGGCGGCCCCCGGATCCCGGCTACGGCGG 102

Job time : 1731 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 16, 2003, 02:39:02 ; Search time 222 Seconds

(without alignments)
1501.332 Million cell updates/sec

Title: US-09-809-545A-2_COPY_1_148

Perfect score: 148

Sequence: 1 MTNKKAVNPYTNMGKLNIPVY.....PATMAAYSDSYGRVYADPY 148

Scoring table:

OLIGO

Xgapop 60.0 , Xgapext 60.0

Ygapop 60.0 , Ygapext 60.0

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 2185239 segs, 112599159 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4368727

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp

-Q=/cgn2_1/USPFO.spool/US09809545/runat_07032003_153954_12150/app_query.faeta.1.327

-DB=N.Geneseq.101002 -QFWT=fastcap -SUFFIX=olip2n.rng -MINMATCH=0.1 -LOOPCL=0

-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi

-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL

-OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000

-USRR=US09809545 @CNC 1.1 396 @runat_07032003_153954_12150 -NCPU=6 -ICPU=3

-NO_XLPEY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120

-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7

-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

N.Geneseq.101002.*

1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*

2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*

3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*

4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*

5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*

6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*

7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*

8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*

9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*

10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*

11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*

12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*

13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*

14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*

15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*

16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*

17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*

18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*

19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*

20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*

21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*

22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.*

23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*

24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	148	100.0	1340	24	AAS94693
2	58	39.2	1513	22	AAH13824
3	58	39.2	1800	22	AAK52245
4	57	38.5	2372	21	AAA07075
5	44	29.7	2118	23	AAS93634
6	38	25.7	481	23	AAS93633
7	35	23.6	539	24	ABK34530
8	32	21.6	607	22	AAH09205
9	20	13.5	1164	22	AAK53229
10	17	11.5	567	23	AAS66245
11	17	11.5	918	22	AAI59249
12	17	11.5	939	22	AAI61034
13	17	11.5	939	22	AAI61035
14	17	11.5	1011	22	AAI59248
15	13	8.8	578	23	AAS90506
16	13	8.8	1252	23	AAS90510
17	13	8.8	1506	21	AAZ46827
18	12	8.1	428	24	ABE68917
19	11	7.4	406	23	AAS93632
20	11	7.4	433	21	AAO04077
21	10	6.8	792	23	AAS67792
22	8	5.4	30	16	AAO97261
23	8	5.4	31	22	AAI29674
24	8	5.4	59	18	AAI773794
25	8	5.4	246	24	ABL39995
26	8	5.4	246	24	ABL40028
27	8	5.4	269	21	AAO08930
28	8	5.4	316	14	AAO60335
29	8	5.4	321	24	ABL40012
30	8	5.4	334	24	ABN21300
31	8	5.4	352	22	ABA82665
32	8	5.4	352	24	ABK22823
33	8	5.4	368	23	AAS57383
34	8	5.4	389	22	AAI55310
35	8	5.4	389	22	ABE52925
36	8	5.4	402	24	ABN79274
37	8	5.4	411	22	AAI29109
38	8	5.4	438	21	AAO47655
39	8	5.4	447	23	AAO30674
40	8	5.4	463	21	ABV03193
41	8	5.4	464	22	ABA58203
42	8	5.4	464	22	AAK06286
43	8	5.4	464	22	AAK31949
44	8	5.4	464	22	AAI37814
45	8	5.4	464	24	ABS06712

ALIGNMENTS

RESULT 1

AAS94693 standard; CDNA: 1340 BP.

AAS94693:

12-MAR-2002 (first entry)

DE Rat secreted factor DNA clone P0184_D11 #1.

XX Rat; secreted factor polypeptide; cardiac disease; renal disease; kidney;

KW inflammatory disease; congestive heart failure; myocarditis; asthma; ss;

KW dilated congestive cardiomyopathy; angina pectoris; cardiac arrhythmia;

KW myocardial infarction; pulmonary hypertension; arteriosclerosis; stroke;

KW atherosclerosis; cardiac tumour; glomerulonephritis; nephrotic syndrome;

KW renal infarction; hereditary nephritis; polycystic kidney disease;

KW chronic renal failure; renal vein thrombosis; medullary sponge kidney;

KW rheumatoid arthritis; osteoarthritis; psoriasis; restenosis; PCR primer;

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

KW	griif versus host reaction; Crohn's disease; ulcerative colitis; probe;		
KV	Alzheimer's disease; gene therapy.		
XX			
OS	Ractus norvegicus.		
XX			
PN	MO200174901-A2.		
XX			
PD	11-OCT-2001.		
XX			
PF	23-MAR-2001; 2001WO-US09555.		
XX			
PR	31-MAR-2000; 2000US-193548P.		
XX			
PR	14-MAR-2001; 2001US-0809545.		
XX			
PA	(SCIO-) SCTOS INC.		
XX			
PI	Stanton LW, White RT;		
XX			
DR	WPI; 2002-010779/01.		
XX			
DR	P-PSDB; AAU70146.		
XX			
PT	Novel secreted factor polypeptide useful for treating cardiac diseases		
XX	such as arteriosclerosis, myocardial infarction, inflammatory diseases		
XX	PT such as asthma, stroke, and rheumatoid arthritis and renal diseases -		
PS	Claim 1; Fig 1; 189pp; English.		
XX			
CC	The invention relates to rat secreted factor polypeptides and the		
XX	polynucleotides encoding them. The sequences are useful for treating		
CC	cardiac, renal or inflammatory diseases. These include cardiac diseases		
CC	such as congestive heart failure, myocarditis, dilated congestive		
CC	cardiomyopathy, angina pectoris, myocardial infarction, cardiac		
CC	arrhythmia, pulmonary hypertension, arteriosclerosis, atherosclerosis and		
CC	cardiac tumours, renal diseases such as glomerulonephritis, nephrotic		
CC	syndrome, renal infarction, hereditary nephritis, polycystic kidney		
CC	disease, chronic renal failure, renal vein thrombosis and medullary		
CC	sponge kidney and inflammatory diseases such as asthma, rheumatoid		
CC	arthritis, osteoarthritis, stroke, psoriasis, reestenosis, graft versus		
CC	host reaction, Crohn's disease, ulcerative colitis and Alzheimer's		
CC	disease. Sequences AA594693-AA594745 represent cDNA clones, which encode		
CC	the secreted factor polypeptides of the invention, and oligonucleotide		
XX	probes and PCR primers.		
XX			
SO	Sequence 1340 BP; 366 A; 394 C; 317 G; 263 T; 0 other;		
Alignment Scores:			
Pred. No.:	2,34e-127	Length:	1340
Score:	148.00	Matches:	148
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	24	Gaps:	0
US-09-809-545A-2_COPY_1_148 (1-148) x AA594693 (1-1340)			
OY	1 MetThrAsnLysLysAlaValaAenProTyrThrAsnGlyTrrLysLsuenProValaI 20		
Db	535 ATGACTAAATAAAAGGCGCTGAACCCCTACACCAATGGCTGGAATTAATCCAGTTGTG 594		
OY	21 GlyAlaValaTyrSerProAspPheTyrAlaGlyThrValleuLeuCyGlnAlaAsnGln 40		
Db	595 GGGCGCGGTACAGCCCGCAGCTTCTATGACGGACGGGTGTGTGGCCAGGCCAACAG 654		
OY	41 GluGlySerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60		
Db	655 GAGGAGATCTTCCATGTACAGTGGCCCGCCAGTCTCTGTATATACTTTCGCAATGCTCGGC 714		
OY	61 PheProTyrProAlaIaIaThrAlaAlaAlaAlaTyrArgLysAlaHisLeuArgLys 80		
Db	715 TTTCATATCCGGCCGCACTGTCGACGTCGATACCGAGGGGCTCATTTCAGGCGGT 774		
OY	81 GlyArgThrValaTyrAnthrPheArgAlaAlaAlaAlaProProProPolleProAlaTyr 100		

Db	775	GGTGCACCGGTACACACACCTTCAGAGCTGGCGCCGCCACCCCAATCCCGGCTAT	834
Qy	101	GLVGLYVALVALTYRGINGLUPROVALTYRGLYASRLYSEULEUNGIMLVLYLYRALA	120
Db	835	GGCGGAGTAGTATCAAGAGCCAGGTGTAGGCAATTAATTCGTACAGGGTGTACGCT	894
Qy	121	ALATYRAGTYRVALAGINPROTPROALATHALALALATYRSEASPSERTYRGLY	140
Db	895	GCAATACCGCTACGGCCCGACCCCTCGCACCTGCTGCTTACAGTACAGTTACGGA	954
Qy	141	ArgValTYRAlaAlaAspProTYR	148
Db	955	CGAGTTATGCTGCCGACCCCTAC	978
RESULT 2			
AAH13824	AAH13824	standard; cDNA; 1513 BP.	
XX	AAH13824;		
XX	AAH13824;		
XX	26-JUN-2001	(first entry)	
XX	Human cDNA sequence SEQ ID NO:10786.		
XX	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss		
XX	Homo sapiens.		
XX	EP1074617-A2.		
XX	07-FEB-2001.		
XX	28-JUL-2000; 2000EP-0116126.		
XX	29-JUL-1999; 99JP-0248036.		
XX	27-AUG-1999; 99JP-0300253.		
XX	11-JAN-2000; 2000JP-0118776.		
XX	02-MAY-2000; 2000JP-0183767.		
XX	09-JUN-2000; 2000JP-0241899.		
XX	(HELI-) HELIX RES INST.		
XX	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
XX	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
XX	WPI; 2001-318749/34.		
XX	Primer sets for synthesizing polynucleotides, particularly the 5602		
XX	full-length cDNAs defined in the specification, and for the detection		
XX	and/or diagnosis of the abnormality of the proteins encoded by the		
XX	full-length cDNAs -		
XX	Claim 8; SEQ ID 10786; 2537bp + CD ROM; English.		
XX	The present invention describes primer sets for synthesizing 5602		
XX	full-length cDNAs defined in the specification. Where a primer set		
XX	comprises: (a) an oligo-dT primer and an oligonucleotide complementary		
XX	to the complementary strand of a polynucleotide which comprises one of		
XX	the 5602 nucleotide sequences defined in the specification, where the		
XX	oligonucleotide comprises at least 15 nucleotides; or (b) a combination		
XX	of an oligonucleotide comprising a sequence complementary to the		
XX	complementary strand of a polynucleotide which comprises a 5'-end		
XX	sequence and an oligonucleotide comprising a sequence complementary to a		
XX	polynucleotide which comprises a 3'-end sequence, where the		
XX	oligonucleotide comprises at least 15 nucleotides and the combination of		
XX	the 5'-end sequence/3'-end sequence is selected from those defined in		
XX	the specification. The primer sets can be used in antisense therapy and		
XX	in gene therapy. The primers are useful for synthesizing polynucleotides		
XX	particularly full-length cDNAs. The primers are also useful for the		
XX	detection and/or diagnosis of the abnormality of the proteins encoded by		
XX	the full-length cDNAs. The primers allow obtaining of the full-length		
XX	cDNAs easily without any specialised methods. AAH0166 to AAH13628 and		
XX	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to		

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 1513 BP; 354 A; 418 C; 395 G; 346 T; 0 other;
 Alignment Scores:
 Pred. No.: 3.57e-44 Length: 1513
 Score: 58.00 Matches: 58
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 39.19% Indels: 0
 DB: 22 Gaps: 0
 US-09-809-545A-2_COPY_1_148 (1-148) x AAH13824 (1-1513)
 QY 49 ProSerSerLeuValTyrThrSerAlaMetProGlyPheProTyrProAlaAlaThrAla 68
 DB 950 CCCAGTTCACCTGTATATACCTTGTGCAATGCCAGGCTTCCGTATCCAGCCACCGCC 1009
 QY 69 AAlaAlaATyrArgGlyAlaHisLeuArgGlyArgGlyValTyrAsnThrPhe 88
 DB 1010 GCGGCGGCTACCGAGGGGCGACCTGCGAGGCCCGGCTCGACCGGTACACCTTC 1069
 QY 89 ArgAlaAlaAlaProProProProPheProAlaTyrGlyGlyValTyrGln 106
 DB 1070 AGGGCGGCGGCGGCGGCGGCGGCTACCGGCTACGGGGTGTGTTCACAG 1123
 RESULT 3
 AAK52245
 ID AAK52245 standard; cDNA; 1800 BP.
 XX
 AC AAK52245;
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 790.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US04098.
 XX
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX
 DR WPI: 2001-476283/51.
 DR P-PSDB: AAM79112.
 XX
 XX Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX
 PS Claim 1; Page 2643-2645; 6221pp; English.

XX
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78333-AAM80302) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 XX
 SQ Sequence 1800 BP; 432 A; 552 C; 481 G; 335 T; 0 other;
 Alignment Scores:
 Pred. No.: 4.18e-44 Length: 1800
 Score: 58.00 Matches: 58
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 39.19% Indels: 0
 DB: 22 Gaps: 0
 US-09-809-545A-2_COPY_1_148 (1-148) x AAK52245 (1-1800)
 QY 49 ProSerSerLeuValTyrThrSerAlaMetProGlyPheProTyrProAlaAlaThrAla 68
 DB 1237 CCCAGTTCACCTGTATATACCTTGTGCAATGCCAGGCTTCCGTATCCAGCCACCGCC 1296
 QY 69 AAlaAlaATyrArgGlyAlaHisLeuArgGlyArgGlyValTyrAsnThrPhe 88
 DB 1297 GCGGCGGCTACCGAGGGGCGACCTGCGAGGCCCGGCTCGACCGGTACACACCTTC 1356
 QY 89 ArgAlaAlaAlaProProProProPheProAlaTyrGlyGlyValTyrGln 106
 DB 1357 AGGGCGGCGGCGGCGGCGGCGGCTACCGGCTACGGGGTGTGTTCACAG 1410
 RESULT 4
 AAA07075
 ID AAA07075 standard; cDNA; 2372 BP.
 XX
 AC AAA07075;
 DT 03-JUL-2000 (first entry)
 XX
 DE cDNA encoding human ataxin-2 binding protein (A2BP).
 XX
 KW Ataxin-2 binding protein; A2BP; human; RNA-binding; cell polarisation;
 KW neuronal plasticity; cellular degeneration signal transduction pathway;
 KW selective RNA transport; spinocerebellar ataxia type-2;
 KW hyperproliferative disorder; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 987..1979
 FT /tag= a
 FT /product= "Human A2BP"
 XX
 PN WO2000012710-A1.
 XX
 PD 09-MAR-2000.
 XX
 PF 01-SEP-1999; 99WO-US20156.
 XX
 PR 01-SEP-1998; 98US-0145391.
 XX
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.
 XX
 PI Pulst SM, Shibata H;

XX WPI, 2000-237873/20.
DR P-PSDB; AAY81462.
XX
XX Nucleic acids encoding an ataxin-2 binding protein useful for
PT inhibiting the expression of active proteins from the SCA2 gene for the
PT treatment of spinocerebellar ataxia type-2.
XX
XX
PS Claim 6; Page 74-77; 82pp; English.

CC This sequence represents cDNA encoding human ataxin-2 binding protein
CC (A2BP). Nucleotide sequences encoding human A2BP were originally
CC isolated in an adult brain cDNA library using the yeast two hybrid
CC method. The ligand of A2BP, ataxin-2, is a 40 kD protein of unknown
CC function that is encoded by the SCA2 gene located on chromosome 12. SCA2
CC has been linked to the autosomal dominant neurodegenerative disorder
CC spinocerebellar ataxia type-2. Individuals afflicted with the disease
CC exhibit CAG triplet expansion in the SCA2 gene, resulting in ataxin-2
CC containing a polyglutamine stretch of about 35-39 residues, whereas that
CC of normal individuals contains approximately 22 contiguous glutamine
CC residues. A2BP and ataxin-2 are components of a cellular degeneration
CC signal transduction pathway. The pathogenic expanded form of ataxin-2 has
CC a higher affinity for A2BP relative to normal ataxin-2; the presence of
CC the expanded form is likely to promote degeneration. A2BP and ataxin have
CC also been found to have a role in gene regulation. The binding of A2BP to
CC ataxin-2 plays an important role in controlling gene expression via the
CC targeting of transport of specific RNAs, selective RNA transport being
CC mediated via the RNA binding domains of A2BP. A2BP is expressed very
CC early in embryonic development. Both ataxin-2 and A2BP are able to bind
CC RNA, and are essential components of the RNA localization network that
CC establishes cellular polarity in embryogenesis. In highly differentiated,
CC polarized cells such as neurons, A2BP and ataxin-2 have a similar
CC function and are required for neuronal plasticity. A2BP nucleic acids may
CC be used for the recombinant production of A2BP proteins or fragments
CC thereof according to standard methodologies. For example, an A2BP protein
CC with an ataxin-2 or RNA binding capability but no signal transduction
CC function can be used as a dominant negative inhibitor of the cellular
CC degeneration signal transduction pathway. A2BP proteins with a signal
CC transduction function can be used to treat hyperproliferative disorders
CC (e.g., cancer) via stimulation of the cellular degeneration pathway.
XX

XX Sequence 2372 BP; 548 A; 726 C; 646 G; 452 T; 0 other;

Alignment Scores:

Pred. No.: 4,52e-43 Length: 2372
Score: 57.00 Matches: 57
Percent Similarity: 100.00% Conservatve: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 38,514 Indels: 0
DB: 21 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x AAA07075 (1-2372)

QY 49 ProSerSeIeuValTyThrSerAlaMetProGlyPheProTyProAlaAlaThrAla 68
DB 1716 CCAGATTCACTGTATATCTTCTGCATAATGCCAGGCTTCCCGATATCAGAGCCACCGCC 1775
QY 69 AlaAlaAlaTyArgGlyValHisIeuArgGlyArgGlyValThrThrValTyraAnThrPhe 88
DB 1776 GGGGGCGCTTACGAGGGGGGACCTGCGAGGGCGGCTGCGACCGTGTACACACCTTC 1835
QY 89 ArgAlaAlaAlaProProProProIleProAlaTyGlyValValValTyTyr 105
DB 1836 AGGGCGGGCGGGCGGGCGGGCGGATCCGGCTACGGGGCGGTGTTTATAC 1886

RESULT 5
AAS93634
ID AAS93634 standard; cDNA; 2118 BP.

XX AAS93634;
XX
XX 13-FEB-2002 (first entry)
XX

DE DNA encoding novel human diagnostic protein #29438.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX Food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PI Dmanac RT, Liu C, Tang YT;

DR WPI, 2001-639362/73.

XX P-PSDB; ABG29447.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

PS Claim 1; SEQ ID No 29438; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensic, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. AAS64197-AAS94564 represent novel human
XX CC nucleotide coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2118 BP; 557 A; 516 C; 542 G; 503 T; 0 other;

Alignment Scores:

Pred. No.: 4,15e-31 Length: 2118
Score: 44.00 Matches: 44
Percent Similarity: 100.00% Conservatve: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 29,73% Indels: 0
DB: 23 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x AAS93634 (1-2118)

QY 59 ProGlyPheProTyProAlaAlaThrAlaAlaAlaTyArgGlyValHisIeuArg 78
DB 586 CCAGGCTTCCCGTATCCAGACCCCGCGCGGCTTACGAGGGGGGACCTGCGCA 645

QY 79 GlyArgGlyValArgThrValTyAsnThrPheArgAlaAlaAlaProProProIlePro 98
DB 646 GGGCGGGCTGCGACCGTGTACACACCTTCAGGGCGCGGGCGGGCGGGCGGATCCCG 705

QY 99 AlaTyArgIyGly 102

Db 706 GCCTACCGCGGT 717
 RESULT 6
 AAS93633
 ID AAS93633 standard; cDNA; 481 BP.
 AC AAS93633;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #29437.
 XX
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 PN WO200175067-A2.
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 XX 23-AUG-2000; 2000US-0649167.
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR P-PSDB; ABG28446.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 1; SEQ ID NO 29437; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 481 BP; 100 A; 145 C; 123 G; 113 T; 0 other;
 Alignment Scores:
 Pred. No.: 3,73e-26 Length: 481
 Score: 38.00 Matches: 38
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 25.68% Indels: 0
 DB: 23 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x AAS93633 (1-481)
 Qy 69 AlaAlaAlaArgGlyAlaHisLeuArgGlyArgGlyArgThrValTYrAsnThrPhe 88
 Db 1 GCGGCCCGCTACCGAGGGGGCGACCTCGAGCGCGGTGCGACCGGTACACACCTTC 60
 Qy 89 ArgAlaAlaAlaProProProProIleProAlaTYrGlyValValTYrGln 106
 Db 61 AGGCGCGCGCGCGCGCGCGCGCGATCCGCGCTACGCGCGGTGTATTACAG 114
 RESULT 7
 ID ABR34530 standard; cDNA; 539 BP.
 AC ABR34530;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Human cDNA for novel secreted protein, SEQ ID 299.
 XX
 KM Human; ss; gene; secreted protein; immune deficiency; viral infection;
 KM bacterial infection; fungal infection; autoimmune disorder; burn;
 KM rheumatoid arthritis; multiple sclerosis; autoimmune thyroiditis;
 KM diabetes; allergy; asthma; neurodegenerative disease; Crohn's disease;
 KM Alzheimer's disease; Parkinson's disease; liver fibrosis; tumour;
 KM coagulation disorder; haemophilia; inflammatory disorder; ulcer;
 KM tissue regeneration; wound healing; haematopoiesis; myeloid deficiency;
 KM lymphoid cell deficiency.
 XX
 OS Homo sapiens.
 XX
 PN WO200177290-A2.
 PD 18-OCT-2001.
 XX
 PF 29-MAR-2001; 2001WO-US10295.
 XX
 PR 06-APR-2000; 2000US-194941P.
 XX
 PA (GEMY) GENETICS INST INC.
 XX
 PI Wong GG, Clark HF, Rechelel K, Agostino MJ, Howes SH, Resnick RJ,
 PI Gulukota K, Graham JR;
 XX
 DR WPI: 2002-179323/23.
 XX
 PT Six hundred and twenty five polynucleotides derived from a variety of
 PT human tissue sources which encode secreted proteins, useful for
 PT treating immune deficiencies and disorders such as autoimmune disorders
 -
 XX
 PS Claim 1; Page 153; 339pp; English.
 XX
 CC The invention relates to 625 polynucleotides which have been derived from
 CC a variety of human tissue sources and sequences that hybridise to them.
 CC proteins, their complements and sequences that hybridise to them.
 CC Also included are a vector comprising the polynucleotide, a host cell
 CC transformed with the vector, the proteins encoded by the
 CC polynucleotides, antibodies that bind to the proteins and identification
 CC of modulators of the proteins or the expression of the polynucleotide.
 CC The polynucleotides can be used as probes for the identification
 CC and isolation of full length cDNA and genomic DNA. The polynucleotides
 CC and proteins can also be used as nutritional supplements. The protein
 CC is useful in the treatment of various immune deficiencies and disorders
 CC such as viral infections, bacterial infections, fungal infections,
 CC autoimmune disorders (e.g. rheumatoid arthritis, multiple sclerosis,
 CC autoimmune thyroiditis and diabetes) and allergic reactions and
 CC conditions (e.g. asthma). They are also useful for treating
 CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
 CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
 CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
 CC useful for tissue regeneration, for wound healing and in the treatment
 CC of burns, incisions and ulcers. The proteins are also useful for

CC regulating haematopoiesis, for treating myeloid or lymphoid cell
CC deficiencies. The present sequence is one of the 625 cDNA sequences
CC encoding a secreted protein.
XX
SQ Sequence 539 BP; 132 A; 132 C; 156 G; 119 T; 0 other;
Alignment Scores:
Pred. No.: 2,44e-23 Length: 539
Score: 35.00 Matches: 35
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 23.65% Indels: 0
DB: 24 Gaps: 0
US-09-809-545A-2_COPY_1_148 (1-148) x ABK34530 (1-539)
Qy 72 TYRARGGLAHHIstenuargGLYARGGLYrThValTYrAsrThrPhaArgAlaIa 91
Db 538 TACCGAGGGGGCGACCTGGAGGCGCGGTCGACCGTACACACTTCAGGGCGCG 479
Qy 92 AAlaProProProleleProAlaTYrGLYGLYValValTYrGln 106
Db 478 GGGCCCCCCCCCGATCCCGGCTACGGCGGTGTTCACAG 434
RESULT 8
AAH09205/c
ID AAH09205 standard; cDNA; 607 BP.
XX AC AAH09205;
XX DT 26-JUN-2001 (first entry)
XX DE Human cDNA clone (3'-primer) SEQ ID NO:6040.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX DR WPI; 2001-318749/34.
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX PS Claim 3; SEQ ID 6040; 2537bp + CD ROM; English.
XX CC The present invention describes primer sets for synthesising 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 607 BP; 140 A; 152 C; 179 G; 129 T; 7 other;
Alignment Scores:
Pred. No.: 1.61e-20 Length: 607
Score: 32.00 Matches: 32
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 21.62% Indels: 0
DB: 22 Gaps: 0
US-09-809-545A-2_COPY_1_148 (1-148) x AAH09205 (1-607)
Qy 117 GLYGLYTYrAlaAlaTYrAlaTYrAlaGlnProThrProAlaThrAlaAlaTYrSer 136
Db 363 GGTGGTATGCTGCTACCGTACGCCGCTACCCCTGCGACATCGCGCTCCTACACT 304
Qy 137 AAspSerTYrGLYrAlaValTYrAlaAlaAspProTYr 148
Db 303 GACAGTTACGAGCAGATTATGCTCGCAGCCCTAC 268
RESULT 9
AAK53229/c
ID AAK53229 standard; cDNA; 1164 BP.
XX AC AAK53229;
XX DT 06-NOV-2001 (first entry)
XX DE Human polynucleotide SEQ ID NO 2758.
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation; ss.
XX OS Homo sapiens.
XX PN WO200157190-A2.
XX PD 09-AUG-2001.
XX PF 05-FEB-2001; 2001WO-US04098.
XX PR 03-FEB-2000; 2000US-0496914.
XX PR 27-APR-2000; 2000US-0560875.
XX PR 20-JUN-2000; 2000US-0598075.
XX PR 19-JUL-2000; 2000US-0620325.
XX PR 01-SEP-2000; 2000US-0654936.
XX PR 15-SEP-2000; 2000US-0663561.
XX PR 20-OCT-2000; 2000US-0693325.
XX PR 30-NOV-2000; 2000US-0728422.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YF, Liu C, Dzmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao Q, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
XX DR WPI; 2001-476283/51.

DR P-PSDB; AAM80096.
 XX Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX
 PS Claim 1; Page 4968; 6221pp; English.
 XX
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polypeptides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 CC
 XX
 SQ Sequence 1164 BP; 273 A; 280 C; 325 G; 286 T; 0 other;
 Alignment Scores:
 Pred. No.: 3.54e-09 Length: 1164
 Score: 20.00 Matches: 20
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 13.51% Indels: 0
 DB: 22 Gaps: 0
 US-09-809-545a-2_COPY_1_148 (1-148) x AAK53229 (1-1164)
 Qy 7 ValaenProTyrrThaenGlyTrpLysLeuAnProValaIglYalaIvalTyrrSerPro 26
 Db 254 GTCAACCTTATACAAATGGCTGGAATTCAGTTGGTGGAGCTCAAGTCCC 195
 RESULT 10
 AAS66245/C
 ID AAS66245 standard; cDNA; 567 BP.
 XX
 AC AAS66245;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #2049.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; AAG02058.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

XX
 PS Claim 1; SEQ ID No 2049; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 567 BP; 92 A; 151 C; 169 G; 155 T; 0 other;
 Alignment Scores:
 Pred. No.: 1.08e-06 Length: 567
 Score: 17.00 Matches: 17
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 11.49% Indels: 0
 DB: 23 Gaps: 0
 US-09-809-545a-2_COPY_1_148 (1-148) x AAS66245 (1-567)
 Qy 8 AsnProTyrrThaenGlyTrpLysLeuAnProValaIglYalaIvalTyrr 24
 Db 480 AACCCCTACACCAACGCGCTGAGACTAATCCAGTGGTGGCGGAGTCTAC 430
 RESULT 11
 AAI59249
 ID AAI59249 standard; cDNA; 918 BP.
 XX
 AC AAI59249;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 1452.
 XX
 KW Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
XX P-PSDB; AAM40093.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 1452; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokine activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX Sequence 918 BP; 235 A; 300 C; 235 G; 148 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 1.68e-06 Length: 918
XX Score: 17.00 Matches: 17
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 11.49% Indels: 0
XX DB: 22 Gaps: 0
XX
XX US-09-809-545a-2_COPY_1_148 (1-148) x AA159249 (1-918)
XX
XX QY 8 AaProTyThraSngLYTrrpYsLeuAaNPovaIvaIGIvAlaValTYr 24
XX DB 459 AACCCCTACACCAAGCGCTGGAAGCTAAATCCAGTGTGGGCGCACTTAC 509
XX
XX RESULT 12
XX AA161034/C
XX ID AA161034 standard; cDNA; 939 BP.
XX
XX AA161034;
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 5023.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokine; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.
XX
XX Homo sapiens.
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
XX P-PSDB; AAM41878.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 5023; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokine activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX Sequence 939 BP; 148 A; 248 C; 297 G; 246 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 1.72e-06 Length: 939
XX Score: 17.00 Matches: 17
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 11.49% Indels: 0
XX DB: 22 Gaps: 0
XX
XX US-09-809-545a-2_COPY_1_148 (1-148) x AA161034 (1-939)
XX
XX QY 8 AaProTyThraSngLYTrrpYsLeuAaNPovaIvaIGIvAlaValTYr 24
XX DB 480 AACCCCTACACCAAGCGCTGGAAGCTAAATCCAGTGTGGGCGCACTTAC 430
XX
XX RESULT 13
XX AA161035/C
XX ID AA161035 standard; cDNA; 939 BP.
XX
XX AA161035;
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 5024.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokine; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.

XX OS Homo sapiens.
XX KW MO200153312-A1.
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US34263.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX DR WPI: 2001-442253/47.
XX DR P-PSDB; AAM41879.
XX PT Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -
XX PS Claim 1; SEQ ID NO 5024; 10078bp; English.
XX CC The invention relates to human nucleic acids (AA157798-AA161369) and
XX CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX CC immunosuppressant and cytostatic activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
XX CC localised neuropathies and central nervous system diseases, such as
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX CC utilisation of the activities such as: Immune system suppression,
XX CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX CC assays for receptor activity, arthritis and inflammation, leukaemias and
XX CC C.N.S disorders.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification.
XX SQ Sequence 939 BP; 148 A; 248 C; 297 G; 246 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 1.72e-06 Length: 939
XX Score: 17.00 Matches: 17
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 11.49% Indels: 0
XX DB: 22 Gaps: 0
XX
XX US-09-809-545A-2_COPY_1_148 (1-148) x AA161035 (1-939)
XX QY 8 AaNPpQYrThraSnglYTrpLysLeuAsnProValaIGlYAlaValaTyr 24
XX AA159248
XX ID AA159248 standard; cDNA; 1011 BP.
XX AC AA159248;
XX XX
XX DT 22-OCT-2001 (first entry)
XX XX

DE DE Human polynucleotide SEQ ID NO 1451.
XX XX
XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX KW peripheral nervous system; neuropathy; central nervous system; CNS;
XX KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX KW leukaemia; ss.
XX OS Homo sapiens.
XX OS KW MO200153312-A1.
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US34263.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX DR WPI: 2001-442253/47.
XX DR P-PSDB; AAM40092.
XX PT Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -
XX PS Claim 1; SEQ ID NO 1451; 10078bp; English.
XX CC The invention relates to human nucleic acids (AA157798-AA161369) and
XX CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX CC immunosuppressant and cytostatic activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
XX CC localised neuropathies and central nervous system diseases, such as
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX CC utilisation of the activities such as: Immune system suppression,
XX CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX CC assays for receptor activity, arthritis and inflammation, leukaemias and
XX CC C.N.S disorders.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification.
XX SQ Sequence 1011 BP; 260 A; 314 C; 267 G; 170 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 1.84e-06 Length: 1011
XX Score: 17.00 Matches: 17
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 11.49% Indels: 0
XX DB: 22 Gaps: 0
XX
XX US-09-809-545A-2_COPY_1_148 (1-148) x AA159248 (1-1011)
XX QY 8 AaNPpQYrThraSnglYTrpLysLeuAsnProValaIGlYAlaValaTyr 24
XX AA159248
XX ID AA159248 standard; cDNA; 1011 BP.
XX AC AA159248;
XX XX
XX DT 552 AACCCCTACACCAACGGCTGGAAGCTAATCCAGTGTGGCGGCGACTTAC 602
XX XX

RESULT 15
AAS90506

ID AAS90506 standard; cDNA; 578 BP.

AC AAS90506;

DT 13-FEB-2002 (First entry)

DE DNA encoding novel human diagnostic protein #26310

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

2000

XX

X

XX
:
o
e
l
:
:

XX
:
:
:
:
:

XX

PR 23-AUG-2000; 2000US-0649167

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT,

DR WPI; 2001-639362/73.

2000

PT diagnostics, fore

PT biodiversity -

PS Claim 1; SEQ ID No 26310; 103pp; English

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridization probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantifying a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pat_sequences](http://wipo.int/pub/published_pat_sequences).

SQ Sequence 578 BP; 152 A; 165 C; 146 G; 113 T; 2 others

Alignment Scores:

Score: 13.00

Best Local Similarity: 100.00%

DB: 23

US-09-809-545A-2_COPY_1_148 (1.

Qy 73 ArgG1yAlaH1sLeuArgG1y

Db 154 AGAGGAGCCATTGAGGGCAGAGGGCGACATATAT 1922

Search completed: March 16, 2003, 04:07:45
Job time : 225 secs

Job time : 225 sec

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 16, 2003, 04:00:32 ; Search time 1442 Seconds
(without alignments)
1662.227 Million cell updates/sec

Title: US-09-809-545a-2_COPY_1_148
Perfect score: 148
Sequence: 1 MTRKKAVPYTNMGKLNPPV.....PATMAAYSDSGRVYADPY 148

Scoring table:
OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 segs, 8097743376 residues

Word size: 1

Total number of hits satisfying chosen parameters: 32308016

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xip
-Q=/cgn2_1/USFTO.spool/US09809545/runat_07032003_153955_12169/app_query.fasta_1.327
-DB=EST -QMT=fastcap -SUFFIX=olip2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALLIGN=400 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pco
-NORM=ext -HEAPSIZE=500 -MILEN=0 -MAXLEN=200000000
-USR=US09809545 @CGN 1.1 2874 @runat_07032003_153955_12169 -NCPU=6 -ICPU=3
-NO_XLPRY -NO_MMAL -LARGEQUERY -NBS_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WALT TIMEOUT=30 -THREDS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlun:*
5: em_estlov:*
6: em_estlpl:*
7: em_estlro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gsa:*
18: em_gsa_hum:*
19: em_gsa_inv:*
20: em_gsa_pln:*
21: em_gsa_vrt:*
22: em_gsa_fun:*
23: em_gsa_mam:*
24: em_gsa_mus:*
25: em_gsa_other:*
26: em_gsa_pro:*
27: em_gsa_yod:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	80	54.1	331	9 A1186273	A1186273 gq20b08.x
2	80	54.1	495	9 A1656926	A1656926 ct48e08.x
3	58	39.2	4039	11 BC026312	BC026312 Homo sapi
4	48	32.4	533	12 BF549922	BF549922 UI-R-E0-b
5	38	25.7	477	10 AW197589	AW197589 XM44G02.x
6	38	25.7	480	9 A1401040	A1401040 ct27412.x
7	38	25.7	773	10 AV729198	AV729198 AV729198
8	38	25.7	773	10 AV729057	AV729057 AV729057
9	36	24.3	439	9 A1968176	A1968176 w14a06.x
10	36	24.3	439	10 AW589795	AW589795 hg22c04.x
11	36	24.3	439	10 BE501380	BE501380 7a41n09.x
12	35	23.6	261	14 F21037	F21037 HSPD05503.H
13	35	23.6	383	9 A1189407	A1189407 gq01h05.x
14	34	23.0	589	17 AC938012	AC938012 NB6-793C
15	32	21.6	218	9 A1189257	A1189257 qc99h03.x
16	32	21.6	367	9 A1095813	A1095813 qb20911.x
17	32	21.6	372	9 A1799929	A1799929 wc41a05.x
18	32	21.6	420	9 A1655094	A1655094 wb67a10.x
19	32	21.6	447	12 BF223478	BF223478 7g33e01.x
20	32	21.6	449	9 A1374891	A1374891 ta60912.x
21	32	21.6	466	9 A1917298	A1917298 t896909.x
22	32	21.6	459	9 A1244212	A1244212 q186b11.x
23	31	20.9	607	9 A1450000	A1450000 AU145000
24	31	20.9	392	9 AA975235	AA975235 qc36c08.s
25	29	19.6	87	9 A1240005	A1240005 qb13906.x
26	28	18.9	408	12 BC795922	BC795922 TUSM SM38
27	28	18.9	521	13 BM183349	BM183349 t106003.x
28	25	16.9	443	9 AA773715	AA773715 a181c05.x
29	25	16.9	733	12 BC306387	BC306387 Em8d11.x
30	23	15.5	655	14 BQ188308	BQ188308 UI-E-EJ1-
31	22	14.9	449	9 A1335996	A1335996 qc42e01.x
32	22	14.9	451	9 A1139843	A1139843 qc72h11.x
33	21	14.2	423	14 R88535	R88535 ym91912.s1
34	21	14.2	553	9 AA496047	AA496047 zv72e08.s
35	20	13.5	291	14 Z25293	Z25293 HSB8B042.S
36	20	13.5	293	9 AA621734	AA621734 a154h09.s
37	20	13.5	302	14 BM930547	BM930547 UI-E-EJ1-
38	20	13.5	304	9 A1962834	A1962834 wq43f08.x
39	20	13.5	306	9 AA405518	AA405518 zw36f08.b
40	20	13.5	323	10 AW295337	AW295337 UI-H-B12-
41	20	13.5	347	9 A1041587	A1041587 ox56g05.x
42	20	13.5	428	9 AA460940	AA460940 zx61d04.s
43	20	13.5	428	10 AV665024	AV665024 AV665024
44	20	13.5	434	13 BM031289	BM031289 4a6653 MA
45	20	13.5	457	9 A1193117	A1193117 qee9h03.x

ALIGNMENTS

RESULT 1
LOCUS A1186273
DEFINITION gq20b08.x1 Soares placenta 8c9weeks 2bhp8tc9w Homo sapiens cDNA
clone IMAGE:1724247 3', mRNA sequence.
ACCESSION A1186273
VERSION A1186273.1 GI:3736911
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 391)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE

TITLE Direct Submission
JOURNAL Submitted (02-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shitaki Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www.shgc.stanford.edu>
 Contact: (Dickson, Mark) mcd@pxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R..M.

FEATURES
 source
 1. 4039
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4815500"
 /tissue_type="Brain, hippocampus"
 /clone_id="NIH MGC_95"
 /lab_host="DH10B"
 /note="Vector: pBluescript"

BASE COUNT 1201 a 858 c 876 g 1104 t

ORIGIN

Alignment Scores:

Pred. No.:	2,23e-43	Length:	4039
Score:	58.00	Matches:	58
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	39.19%	Indels:	0
DB:	11	Gaps:	0

US-09-809-545a-2_COPY_1_148 (1-148) x BC026312 (1-4039)

Qy 49 ProSerSerLeuValTYrThSerAlaMetProGlyPheProTYrProAlaAlaThAla 68
 |||||
 Db 967 CCGAGTTCACTTGATATACCTTCTGCATGCGAGCTTCCGATACGAGCAGCCGCC 1026

Qy 69 AAlaAlaAlaTYrAGGlyAlaHisLeuArgGlyTYrAGGlyThValTYrAsnThrPhe 88
 |||||
 Db 1027 GCGGCGCGCTTCCGAGGCGCAGCTTGCAGGCGCGGTGCACCGTGTACACACCTTC 1086

Qy 89 ArgAlaAlaAlaProProProProProAlaTYrGlyValValTYrGln 106
 |||||
 Db 1087 AGGGCCGCGCGCGCGCGCGCGCGCGCGCGCTTCCGCGGCGGTGTTTACCAAG 1140

RESULT 4
LOCUS BF549922 533 bp mRNA linear EST 12-DEC-2000
DEFINITION UI-R-80-bv-c-04-0-UI.r1 UI-R-80 Rattus norvegicus cDNA clone
 UI-R-80-bv-c-04-0-UI 5', mRNA sequence.
ACCESSION BF549922
VERSION BF549922.1 GI:11659652
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
REFERENCE 1 (bases 1 to 533)

AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: meares@blue.weeg.uiowa.edu
 cDNA Library Preparation: M.B. Soares Lab Clone Distribution:
 Clones will be available through Research Genetics (www.resgen.com)
 This clone is also available through the I.M.A.G.E. Consortium at LNL (info@image.lnl.gov). IMAGE ID= 1777166
 Seq primer: M13 Forward

FEATURES
 source
 1. 533
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-80-bv-c-04-0-UI"
 /clone_id="UI-R-80"
 /dev_stage="embryonic"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: p773D-Pac (Pharmacia) with a modified polylinker. Site 1: NotI; Site 2: EcoRI. This library consists of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."

BASE COUNT 134 a 125 c 164 g 109 t 1 others

ORIGIN

Alignment Scores:

Pred. No.:	8.44e-35	Length:	533
Score:	48.00 <td>Matches:</td> <td>48</td>	Matches:	48
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	32.43%	Indels:	0
DB:	12	Gaps:	0

US-09-809-545a-2_COPY_1_148 (1-148) x BF549922 (1-533)

Qy 59 ProGlyPheProTYrProAlaAlaThAlaAlaTYrAGGlyAlaHisLeuArg 78
 |||||
 Db 460 CCGGCTTCCATATCCGCGCGCCAGCTGCATGACGAGGCGCTCACCTTCA 401

Qy 79 GLyArgGlyArgThrValTYrAsnThrPheArgAlaAlaProProProProLeu 98
 |||||
 Db 400 GCGGTGCTCCACCGGTGTACACCTTCAGACTGCGGCGCGCCGCCCAATCCG 341

Qy 99 AAlaTYrGlyGlyValValTYrGln 106
 |||||
 Db 340 GCCTATGCGGCGGTGTTTACCAAG 317

RESULT 5
LOCUS AM197589 477 bp mRNA linear EST 29-NOV-1999
DEFINITION XM44602.x1 NCI CGAP GC6 Homo sapiens cDNA clone IMAGE:2687090 3' similar to TR:043251 043251 HYPOTHETICAL 39.5 KD PROTEIN. ;, mRNA sequence.
ACCESSION AM197589
VERSION AM197589.1 GI:6476819
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 477)

ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 599)
AUTHORS	Zabarovskiy,E.R., Glazutulin,R., Podowski,R.M., Zabarovska,V.V., Xie L., Muravenko,O.V., Kozirev,S., Petrenko,L., Skobeleva,N., Li,J., Pitropov,A., Kashba,V., Erberg,I., Winberg,G. and Mahlestedt,C.
TITLE	Nctf clones in the analysis of the human genome
JOURNAL	Nucleic Acids Res. 28 (7), 1635-1639 (2000)
MEDLINE	20175728
COMMENT	Contact: Podowski RM Center for Genomics Research Karolinska Institute 17177 Stockholm Sweden Tel.: +46-8-728-6372 Fax: +46-8-337983 Email: Raf.Podowski@cgr.ki.se Class: Nctf site.
FEATURES	Location/Qualifiers 1..599 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="Human Nctf clones" /note="Organ: Lung; DNA was isolated from A549 cells after sodium arsenite exposure for 4 weeks. This fragment was differentially methylated relative to untreated controls and was identified using methylation sensitive AP-PCR and sequenced."
BASE COUNT	131 a 171 c 157 g 140 t
ORIGIN	
Alignment Scores:	
Pred. No.:	9.8e-22 Length: 599
Score:	34.00 Matches: 34
Percent Similarity:	100.00% Conservativity: 0
Best local Similarity:	100.00% Mismatches: 0
Query Match:	22.97% Indels: 0
Dbl:	17 Gaps: 0
US-09-809-545A-2_COPY_1_148 (1-148) x AQ938012 (1-599)	
Cy	69 AATAAATATYARGGIYAIAHSIEUAAGTGYAARGIYAAGThValTYRanThrPhe 88
Dbl	1 GGCGCCGCTACCGAGGGCGCACCCTGCCAGCGCCGGCTGCACCTGTAAACAACCTTC 60
Cy	89 ATGAATAAATAAPROBPROBPROBIOLEPROAATYATYGIYGI 102
Dbl	61 AGCGCCGCGCGCCCCCGCCCGAATCCCGGCTACCGCGGG 102
RESULT 15	
LOCUS	AII89257 218 bp mRNA linear EST 28-OCT-1998
DEFINITION	qc99h03.xl Soares_pregnant uterus_NbHPU Homo sapiens CDNA clone
ACCESSION	AII89257 IMAGE:1722389 3', mRNA sequence.
VERSION	AII89257
KEYWORDS	AII89257.1 GI:3740466
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 218) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-rt@mail.nih.gov This clone is available royally-free through LBL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 484 Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence ecop.201.

FEATURES	location/Qualifiers
source	1. 218
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="IMAGE:1722389"
	/clone_1b="Soares_pregnant_uterus_NbHPu"
	/sex="female"
	/dev_stage="adult"
	/lab_host="DH10B"
	/note="Organ: uterus; Vector: pT7T3-Pac; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo (dT) primer [5' AACTCGAAGATTTCGGCGCCCTTTTTTTTTTTT 3'] , double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia) , digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."
BASE COUNT	41 a 70 c 54 g 53 t
ORIGIN	
Alignment Scores:	
Pred. No.:	3,096/20
Score:	32.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	21.62%
DB:	9 Gaps: 0
US-809-545A-2_COPY_1_148 (1-148) x A1189257 (1-218)	
QY	117 GYGYLYRRLAALATYRATYRAGINDPROTHPRQALATRHALALATYRISer 136
DB	17 GGTGGTATGCTGCATACCGCTACGCCACAGCCTACCCCTGCACATCGCGCTGCCTACAGT 76
QY	137 AepserTYRGLYARGVALTYRRLAALASpProTYr 148
DB	77 GACAGTACGACACAGTTATGCTGCCAGACCCCTTAC 112

Search completed: March 16, 2003, 05:00:41
Job time : 1445 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 16, 2003, 04:02:33 ; Search time 74 Seconds
(without alignments)
613.354 Million cell updates/sec

Title: US-09-809-545A-2_COPY_1_148
Perfect score: 148
Sequence: 1 MTRKKAHPYTNQMKLNPVY.....PATMAAYSDEYGRVYADPY 148

Scoring table:

	OLIGO	Xgapop 60.0	Xgapext 60.0
		Xgapop 60.0	Xgapext 60.0
		Fgapop 6.0	Fgapext 7.0
		Delop 6.0	Delext 7.0

Searched: 441362 seqs, 153338381 residues

Word size: 1

Total number of hits satisfying chosen parameters: 878600

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=x1p
-Q=/cgn2_1/USFTPO.spool/US09809545/runat.07032003.153955.12186/app.query.fasta.1.327
-DB=Issued_Patents_NA -OPMT=fastap -SUFFIX=olip2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09809545 @cgn 1.1 61 @runat.07032003.153955.12186 -NCPU=6 -ICPU=3
-NO_XLPRY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57	38.5	2372	4	US-09-145-391-1
2	13	8.8	1506	3	US-09-176-657-5
3	8	5.4	1007	3	US-09-248-335-31
4	5	5.4	1279	3	US-09-248-335-25
5	8	5.4	1416	3	US-08-866-928B-2
6	7	8	1568	4	US-09-043-937A-1
7	8	5.4	2167	4	US-09-489-869-10
8	8	5.4	2249	3	US-08-860-635A-18
9	8	5.4	2249	3	US-09-281-476-18
10	8	5.4	3923	4	US-08-860-635A-20
11	8	5.4	3923	4	US-09-281-476-20
12	8	5.4	4060	1	US-08-308-949A-1

13	8	5.4	5357	4	US-09-392-184-5	Sequence 5, Appli
14	8	5.4	4403765	4	US-09-103-840A-2	Sequence 2, Appli
15	8	5.4	4403765	4	US-09-103-840A-2	Sequence 2, Appli
16	8	5.4	4411529	4	US-09-103-840A-1	Sequence 1, Appli
17	8	5.4	4411529	4	US-09-103-840A-1	Sequence 1, Appli
18	7	4.7	24	1	US-08-411-796-187	Sequence 187, App
19	7	4.7	24	3	US-08-471-039-187	Sequence 187, App
20	7	4.7	24	3	PCT-US93-11198-187	Sequence 187, App
21	7	4.7	25	6	5504194-3	Patent No. 5504194
22	7	4.7	30	1	US-08-361-920-76	Sequence 76, Appli
23	7	4.7	30	1	US-08-479-939-76	Sequence 76, Appli
24	7	4.7	30	1	US-08-483-432-76	Sequence 76, Appli
25	7	4.7	30	5	PCT-US95-07372-3	Sequence 3, Appli
26	7	4.7	42	4	US-09-162-484-7	Sequence 7, Appli
27	7	4.7	51	2	US-08-350-260A-565	Sequence 565, App
28	7	4.7	53	1	US-08-081-539-59	Sequence 59, Appli
29	7	4.7	53	1	US-08-466-647-59	Sequence 59, Appli
30	7	4.7	54	1	US-07-981-651-3	Sequence 3, Appli
31	7	4.7	54	1	US-08-279-996-3	Sequence 3, Appli
32	7	4.7	57	2	US-07-814-220-26	Sequence 26, Appli
33	7	4.7	57	2	US-07-812-421-26	Sequence 26, Appli
34	7	4.7	57	4	US-09-043-303-14	Sequence 14, Appli
35	7	4.7	61	1	US-08-081-539-58	Sequence 58, Appli
36	7	4.7	61	1	US-08-466-647-58	Sequence 58, Appli
37	7	4.7	62	2	US-07-814-220-25	Sequence 25, Appli
38	7	4.7	62	2	US-07-812-421-25	Sequence 25, Appli
39	7	4.7	64	2	US-07-814-220-22	Sequence 22, Appli
40	7	4.7	64	2	US-07-812-421-22	Sequence 22, Appli
41	7	4.7	65	1	US-08-411-796-84	Sequence 84, Appli
42	7	4.7	65	1	US-08-411-796-85	Sequence 85, Appli
43	7	4.7	65	1	US-08-411-796-86	Sequence 86, Appli
44	7	4.7	65	1	US-08-411-796-87	Sequence 87, Appli
45	7	4.7	65	1	US-08-411-796-88	Sequence 88, Appli

ALIGNMENTS

RESULT 1
US-09-145-391-1
Sequence 1, Application US/09145391
Patent No. 6194171
GENERAL INFORMATION:
APPLICANT: Pulist, Stefan M.
APPLICANT: Shubata, Hiroki
TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
TITLE OF INVENTION: Products Related Thereto and Methods of Using Same
FILE REFERENCE: CE 3093
CURRENT APPLICATION NUMBER: US/09/145.391
CURRENT FILING DATE: 1998-09-01
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 2372
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (987)..(1979)
US-09-145-391-1

Alignment Scores:
Pred. No.: 2,39e-42
Score: 57.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 38.51%
DB: 4
Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-145-391-1 (1-2372)

OY 49 PROSERGIEVALTYTThSERALAMETPROGTYRPRPQALALATHRALA 68
DB 1716 CCAGTTCACCTTGATATATATCTTGCAATGCAGGCTTCCTATCCAGCAGCCGCC 1775

Qy 69 AlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArgThrValTyrIleThrPhe 88
Db 1776 GCGGCGCGCTACGAGGGGGGACCTGCGAGCGCGGTGCGACCTGACCAACACCTTC 1835

Qy 89 ArgAlaAlaAlaProProProProlleProAlaTyrGlyValValTyr 105
Db 1836 AGGGCGCGCGCGCGCGCGCGCGCGATCCGGGCTACGGCGGTGTGTTC 1886

RESULT 2

US-09-176-657-5
; Sequence 5, Application US/09176657
; Patent No. 6020164
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Tang, Y. Tom
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Lu, Aina
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: HUMAN RNA BINDING PROTEINS
; FILE REFERENCE: PF-0611 US
; CURRENT APPLICATION NUMBER: US/09/176,657
; EARLIER FILING DATE: 1998-10-21
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 1506
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 1250374
US-09-176-657-5

Alignment Scores:
Pred. No.: 0.00481 Length: 1506
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.78% Indels: 0
DB: 3 Gaps: 0

US-09-809-545a-2_COPY_1_148 (1-148) x US-09-176-657-5 (1-1506)

Qy 73 ArgGlyAlaHisLeuArgGlyArgThrValTyr 85
Db 999 AGAGAGAGCCCATTTGAGGGGAGAGGGCGGACGATATAT 1037

RESULT 3

US-09-248-335-31
; Sequence 31, Application US/09248335
; Patent No. 6096504
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEF, DANIEL
; TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES
; FILE REFERENCE: CL-1128-A
; CURRENT APPLICATION NUMBER: US/09/248,335
; EARLIER FILING DATE: 1999-02-10
; EARLIER APPLICATION NUMBER: 08/924,759
; EARLIER FILING DATE: 1997-September-05
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 31
; LENGTH: 1007
; TYPE: DNA
; ORGANISM: maize
US-09-248-335-31

Alignment Scores:
Pred. No.: 103 Length: 1007
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 3 Gaps: 0

US-09-809-545a-2_COPY_1_148 (1-148) x US-09-248-335-31 (1-1007)

Qy 73 ArgGlyAlaHisLeuArgGlyArg 80
Db 234 CGAGGCGCTCACCTTCGAGACG 257

RESULT 4

US-09-248-335-25
; Sequence 25, Application US/09248335
; Patent No. 6096504
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEF, DANIEL
; TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES
; FILE REFERENCE: CL-1128-A
; CURRENT APPLICATION NUMBER: US/09/248,335
; EARLIER FILING DATE: 1999-02-10
; EARLIER APPLICATION NUMBER: 08/924,759
; EARLIER FILING DATE: 1997-September-05
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 25
; LENGTH: 1279
; TYPE: DNA
; ORGANISM: maize
US-09-248-335-25

Alignment Scores:
Pred. No.: 128 Length: 1279
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 3 Gaps: 0

US-09-809-545a-2_COPY_1_148 (1-148) x US-09-248-335-25 (1-1279)

Qy 73 ArgGlyAlaHisLeuArgGlyArg 80
Db 196 CGAGGCGCTCACCTTCGAGACGA 219

RESULT 5

US-08-866-928B-2
; Sequence 2, Application US/08866928B
; Patent No. 6159702
; GENERAL INFORMATION:
; APPLICANT: Traish, Abdulmaged M
; TITLE OF INVENTION: In-vivo diagnostic methods for determining
; TITLE OF INVENTION: whether a primary breast tumor is clinically metastatic
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David Prashker, P.C.
; STREET: P.O. Box 5387
; CITY: Magnolia
; STATE: MA
; COUNTRY: USA
; ZIP: 01930
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 Mb storage
; COMPUTER: IBM PS1
; OPERATING SYSTEM: MS DOS
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/866,928B
; FILING DATE: May 30, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 509,570
; FILING DATE: July 31, 1995

ATTORNEY/AGENT INFORMATION:
NAME: David Prashker
REGISTRATION NUMBER: 29,693
REFERENCE/DOCKET NUMBER: BUMC-060
TELECOMMUNICATION INFORMATION:
TELEPHONE: (978) 525-3794
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1416
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-866-928B-2

Alignment Scores:
Pred. No.: 139 Length: 1416
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 3 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-08-866-928B-2 (1-1416)

OY 64 ProalaalathraAlaAlaAla 71
Db 87 CCAGCAGCACGACGACGCCGCC 110

RESULT 6
US-09-043-937A-1
Sequence 1, Application US/09043937A
Patent No. 6211432
GENERAL INFORMATION:
APPLICANT: BOUDET, ALAIN-MICHEL
PICHON, MAGALIE
GRIMA-PETTERATI, JACQUELINE
BECKERT, MICHEL
GAMAS, PASCAL
BRIAT, JEAN-FRANCOIS
TITLE OF INVENTION: DNA SEQUENCES CODING FOR CINNAMOYL-CoA
REDUCTASE, AND APPLICATIONS THEREOF IN THE CONTROL OF
LIGNIN CONTENTS IN PLANTS
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE, P.C.
STREET: 110 NORTH GLEBE ROAD, 8TH FLOOR
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/043,937A
FILING DATE: 24-Jul-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR96/01544
FILING DATE: 03-OCT-1996
APPLICATION NUMBER: FR 95.11623
FILING DATE: 03-OCT-1995
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36,663
REFERENCE/DOCKET NUMBER: 1487-20
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-816-4100
TELEFAX: 703-816-4100
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 1568 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 278..1306
SEQUENCE DESCRIPTION: SEQ ID NO: 1
US-09-043-937A-1

Alignment Scores:
Pred. No.: 153 Length: 1568
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 4 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-043-937A-1 (1-1568)

OY 64 ProalaalathraAlaAlaAla 71
Db 281 CCGCGCGCTACCGCAGCGCGGCC 304

RESULT 7
US-09-489-869-10/c
Sequence 10, Application US/09489869A
Patent No. 6268151
GENERAL INFORMATION:
APPLICANT: Susan Murray
APPLICANT: Lex M. Cowser
TITLE OF INVENTION: ANTISENSE MODULATION OF MACROPHAGE MIGRATION INHIBITORY FACTOR
FILE REFERENCE: RTS-0110
CURRENT APPLICATION NUMBER: US/09/489,869A
CURRENT FILING DATE: 2000-01-20
NUMBER OF SEQ ID NOS: 88
SEQ ID NO 10
LENGTH: 2167
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1173)...(1280)
FEATURE:
NAME/KEY: CDS
LOCATION: (1470)...(1642)
FEATURE:
NAME/KEY: CDS
LOCATION: (1738)...(1804)
US-09-489-869-10

Alignment Scores:
Pred. No.: 203 Length: 2167
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 4 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-489-869-10 (1-2167)

OY 89 ArgAlaAlaAlaProProProPro 96
Db 1684 CGCGCGCGCGCGCGCTCCCGCGGCC 1661

RESULT 8
US-08-860-635A-18/c
Sequence 18, Application US/08860635A
Patent No. 6143878
GENERAL INFORMATION:
APPLICANT: Koopman, Peter

APPLICANT: Goodfellow, Peter
TITLE OF INVENTION: SOX-9 GENE AND PROTEIN AND
TITLE OF INVENTION: USE IN THE REGENERATION OF BONE OR CARTILAGE
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: U.S.A.
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,635A
FILING DATE: 29-MAY-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PM9714
FILING DATE: 29-NOV-1994
APPLICATION NUMBER: AU PM9835
FILING DATE: 05-DEC-1994
APPLICATION NUMBER: PCT/AU95/00799
FILING DATE: 29-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10981
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX:
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2249 base pairs
TYPE: nucleic acid
STRADEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-860-635A-18
Alignment Scores:
Pred. No.: 210 Length: 2249
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
Gaps: 0
DB: 3
US-09-809-545A-2_COPY_1_148 (1-148) x US-08-860-635A-18 (1-2249)
Qy 89 ArgAlaAlaProProProPro 96
Db 788 CCGCAGCGCGCTCCGCTCTCCA 765
RESULT 9
US-09-281-476-18/C
Sequence 18, Application US/09281476
GENERAL INFORMATION:
APPLICANT: Koopman, Peter
APPLICANT: Goodfellow, Peter
TITLE OF INVENTION: SOX-9 GENE AND PROTEIN AND
TITLE OF INVENTION: USE IN THE REGENERATION OF BONE OR CARTILAGE
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY

COUNTRY: U.S.A.
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/281,476
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/860,635
FILING DATE:
APPLICATION NUMBER: AU PM9835
FILING DATE: 05-DEC-1994
APPLICATION NUMBER: PCT/AU95/00799
FILING DATE: 29-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10981
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX:
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2249 base pairs
TYPE: nucleic acid
STRADEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-281-476-18
Alignment Scores:
Pred. No.: 210 Length: 2249
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
Gaps: 0
DB: 4
US-09-809-545A-2_COPY_1_148 (1-148) x US-09-281-476-18 (1-2249)
Qy 89 ArgAlaAlaProProProPro 96
Db 788 CCGCAGCGCGCTCCGCTCTCCA 765
RESULT 10
US-08-860-635A-20/C
Sequence 20, Application US/08860635A
GENERAL INFORMATION:
APPLICANT: Koopman, Peter
APPLICANT: Goodfellow, Peter
TITLE OF INVENTION: SOX-9 GENE AND PROTEIN AND
TITLE OF INVENTION: USE IN THE REGENERATION OF BONE OR CARTILAGE
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: U.S.A.
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA: US/08/860,635A

FILING DATE: 29-MAY-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PM9714
FILING DATE: 29-NOV-1994
APPLICATION NUMBER: AU PM9835
FILING DATE: 05-DEC-1994
APPLICATION NUMBER: PCT/AU95/00799
FILING DATE: 29-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10981
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX:
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 3923 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-860-635A-20

Alignment Scores:
Pred. No.: 342 Length: 3923
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 3 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-08-860-635A-20 (1-3923)

QY 89 ArgAlaAlaIaProProProPro 96
Db 846 CGCGAGCGGCTCGGCTCTCCA 823

RESULT 11
US-09-281-476-20/c
Sequence 20, Application US/09281476
Patent No. 6316597
GENERAL INFORMATION:
APPLICANT: Koopman, Peter
APPLICANT: Goodfellow, Peter
TITLE OF INVENTION: SOX-9 GENE AND PROTEIN AND
TITLE OF INVENTION: USE IN THE REGENERATION OF BONE OR CARTILAGE
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: U.S.A.
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/281.476
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/860, 635
FILING DATE:
APPLICATION NUMBER: AU PM9835
FILING DATE: 05-DEC-1994
APPLICATION NUMBER: PCT/AU95/00799
FILING DATE: 29-NOV-1995

ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10981
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX:
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 3923 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-281-476-20

Alignment Scores:
Pred. No.: 342 Length: 3923
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 4 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-281-476-20 (1-3923)

QY 89 ArgAlaAlaIaProProProPro 96
Db 846 CGCGAGCGGCTCGGCTCTCCA 823

RESULT 12
US-08-308-949A-1/c
Sequence 1, Application US/08308949A
Patent No. 5580703
GENERAL INFORMATION:
APPLICANT: Kotin, Robert M.
APPLICANT: Berns, Kenneth I.
APPLICANT: Linden, Ralph M.
TITLE OF INVENTION: Human Adeno-Associated Virus Integration
TITLE OF INVENTION: Site DNA and Uses Thereof
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/308.949A
FILING DATE: September 20, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/947,127
FILING DATE: September 27, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Alice O.
REGISTRATION NUMBER: 33,542
REFERENCE/DOCKET NUMBER: ACC92-10F
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-9540
TELEFAX: 617-861-6240
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4060 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

```
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
US-08-308-949A-1

Alignment Scores:
Pred. No.: 353 Length: 4060
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 1 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-08-308-949A-1 (1-4060)
Qy 89 ArgAlaAlaAlaProProProPro 96
Db 484 CCGGCAGCGCGCGCGCGCGCGCG 461

RESULT 13
US-09-392-184-5
/ Sequence 5, Application US/09392184
/ Patent No. 6395889
/ GENERAL INFORMATION:
/ APPLICANT: Robison, Keith E.
/ TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN
/ FILE REFERENCE: 5800-55
/ CURRENT APPLICATION NUMBER: US/09/392,184
/ CURRENT FILING DATE: 1999-09-09
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 5
/ LENGTH: 5357
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(5357)
/ OTHER INFORMATION: repolysin (ADAM family of metalloprotease)
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(5357)
/ OTHER INFORMATION: n = A, T, C or G
US-09-392-184-5

Alignment Scores:
Pred. No.: 450 Length: 5357
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 4 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-392-184-5 (1-5357)
Qy 64 ProAlaAlaThAlaAlaAlaAla 71
Db 4638 CCGCGCGCACGCTGCGCGCTGCT 4661

RESULT 14
US-09-103-840A-2
/ Sequence 2, Application US/09103840A
/ Patent No. 6294328
/ GENERAL INFORMATION:
/ APPLICANT: FLEISCHMAN, Robert D.
/ APPLICANT: WHITE, Owen R.
/ APPLICANT: FRASER, Claire M.
/ TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
/ FILE REFERENCE: 24366-20007.00
/ CURRENT APPLICATION NUMBER: US/09/103,840A
/ CURRENT FILING DATE: 1998-06-24
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 4403765
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ FEATURE:
/ OTHER INFORMATION: "n" bases at various positions throughout the sequence
/ OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 1.52e+05 Length: 4403765
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 4 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-103-840A-2 (1-4403765)
Qy 64 ProAlaAlaThAlaAlaAlaAla 71
Db 2932874 CCGGCTGCACGCGCGCGCGCGCGCC 2932851

SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 4403765
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ FEATURE:
/ OTHER INFORMATION: CDC 1551
/ OTHER INFORMATION: "n" bases at various positions throughout the sequence
/ OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 1.52e+05 Length: 4403765
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 4 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-103-840A-2 (1-4403765)
Qy 90 AlaAlaAlaProProProProle 97
Db 2295213 GCCGCTGCACCGCCTCCACCGATC 2295236

RESULT 15
US-09-103-840A-2/C
/ Sequence 2, Application US/09103840A
/ Patent No. 6294328
/ GENERAL INFORMATION:
/ APPLICANT: FLEISCHMAN, Robert D.
/ APPLICANT: WHITE, Owen R.
/ APPLICANT: FRASER, Claire M.
/ APPLICANT: VENTER, John C.
/ TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
/ FILE REFERENCE: 24366-20007.00
/ CURRENT APPLICATION NUMBER: US/09/103,840A
/ CURRENT FILING DATE: 1998-06-24
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 4403765
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ FEATURE:
/ OTHER INFORMATION: CDC 1551
/ OTHER INFORMATION: "n" bases at various positions throughout the sequence
/ OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 1.52e+05 Length: 4403765
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 4 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-103-840A-2 (1-4403765)
Qy 64 ProAlaAlaThAlaAlaAlaAla 71
Db 2932874 CCGGCTGCACGCGCGCGCGCGCGCC 2932851

Search completed: March 16, 2003, 05:21:14
Job time : 1223 secs
```

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 16, 2003, 04:04:07 ; Search time 98 Seconds
(without alignments)
1059,960 Million cell updates/sec

Title: US-09-809-545A-2_COPY_1_148
Perfect score: 148
Sequence: 1 MTRKKAVPYTNMGKLNIPVY.....PATMAAYSDSYGRVYADPY 148

Scoring table:

	OLIGO
Xgapop 60.0 , Xgapext 60.0	
Igapop 60.0 , Igapext 60.0	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 501302 seqs, 350932545 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1001176

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USFTO.spool/US09809545/rnatc_07032003_153955_12214/app_query.fasta.1.327
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=olip2n.rmpb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=plc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09809545 @CCN 1 1.80 @rnatc_07032003_153955_12214
-NCPU=6 -ICPU=3 -NO XLPXT -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:*

1:	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2:	/cgn2_6/ptodata/2/pubpna/PTI_NEW_PUB.seq:*
3:	/cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4:	/cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5:	/cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6:	/cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7:	/cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8:	/cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10:	/cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11:	/cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12:	/cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13:	/cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14:	/cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	148	100.0	1340	US-09-809-545A-1	Sequence 1, Appl1
2	57	38.5	2372	US-09-794-591-1	Sequence 1, Appl1
3	8	5.4	234	US-09-878-574-3090	Sequence 3090, Ap
4	8	5.4	251	US-09-923-876-1816	Sequence 1816, Ap

C 5	8	5.4	263	10	US-09-923-876-4839	Sequence 4839, Ap
C 6	8	5.4	355	10	US-09-983-965-5536	Sequence 5536, Ap
C 7	8	5.4	388	10	US-09-815-343-59	Sequence 59, Appl
C 8	8	5.4	389	10	US-09-864-761-32302	Sequence 32302, A
C 9	8	5.4	407	10	US-09-960-352-8373	Sequence 8373, Ap
C 10	8	5.4	411	9	US-10-025-380-659	Sequence 659, App
C 11	8	5.4	411	10	US-09-922-217-659	Sequence 659, App
C 12	8	5.4	411	10	US-09-833-263-659	Sequence 659, App
C 13	8	5.4	464	10	US-09-864-761-11104	Sequence 11104, A
C 14	8	5.4	501	10	US-09-833-790-175	Sequence 175, App
C 15	8	5.4	587	10	US-09-864-761-15794	Sequence 15794, A
C 16	8	5.4	726	9	US-10-188-947-2	Sequence 2, Appl1
C 17	8	5.4	739	10	US-09-752-639-2	Sequence 2, Appl1
C 18	8	5.4	739	10	US-09-984-198-2	Sequence 2, Appl1
C 19	8	5.4	763	10	US-09-764-877-3667	Sequence 3667, Ap
C 20	8	5.4	1091	10	US-09-822-849A-388	Sequence 388, App
C 21	8	5.4	1225	10	US-09-814-777A-101	Sequence 101, App
C 22	8	5.4	1681	10	US-09-999-254-1	Sequence 1, Appl1
C 23	8	5.4	1902	9	US-10-086-464-10	Sequence 10, Appl1
C 24	8	5.4	1939	9	US-10-086-464-9	Sequence 9, Appl1
C 25	8	5.4	2000	10	US-09-887-576-853	Sequence 853, App
C 26	8	5.4	2034	12	US-10-044-090-209	Sequence 209, App
C 27	8	5.4	2167	10	US-09-969-708-322	Sequence 322, App
C 28	8	5.4	2167	10	US-09-880-107-2171	Sequence 2171, Ap
C 29	8	5.4	2167	12	US-10-029-654-1	Sequence 12, App
C 30	8	5.4	2249	10	US-09-910-087-18	Sequence 18, Appl1
C 31	8	5.4	2307	9	US-09-893-519A-87	Sequence 87, Appl1
C 32	8	5.4	2593	10	US-09-880-107-3278	Sequence 3278, Ap
C 33	8	5.4	3923	10	US-09-910-087-20	Sequence 20, Appl
C 34	8	5.4	3923	10	US-09-880-107-3935	Sequence 3935, Ap
C 35	8	5.4	3923	10	US-09-967-768A-222	Sequence 222, App
C 36	8	5.4	4595	10	US-09-917-800A-1531	Sequence 1531, Ap
C 37	8	5.4	5063	9	US-09-931-375A-1	Sequence 1, Appl1
C 38	8	5.4	6188	9	US-09-736-457-1823	Sequence 1823, Ap
C 39	8	5.4	6188	9	US-09-902-941-1823	Sequence 1823, Ap
C 40	8	5.4	6188	9	US-09-849-626-1823	Sequence 1823, Ap
C 41	8	5.4	6188	10	US-09-833-790-2132	Sequence 2132, App
C 42	8	5.4	6282	9	US-10-108-605-212	Sequence 212, App
C 43	8	5.4	7029	10	US-09-826-191-1	Sequence 1, Appl1
C 44	8	5.4	7090	9	US-09-832-392-28	Sequence 28, Appl
C 45	8	5.4	11190	9	US-10-108-605-266	Sequence 266, App

ALIGNMENTS

RESULT 1
US-09-809-545A-1
; Sequence 1, Application US/09809545A
; Patent No. US20020110804A1
; GENERAL INFORMATION:
; APPLICANT: Stanton, Lawrence W.
; TITLE OF INVENTION: SECRETED FACTORS
; FILE REFERENCE: SCIOS.017A
; CURRENT APPLICATION NUMBER: US/09/809,545A
; CURRENT FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PasteSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1340
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-09-809-545A-1

Alignment Scores:
Pred. No.: 1,28e-133 Length: 1340
Score: 148.00 Matches: 148
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0
US-09-809-545A-2_COPY_1_148 (1-148) x US-09-809-545A-1 (1-1340)

```
Oy 1 MetThrAsnLysLeuValAlaValAsnProTyrThrAsnGlyTrrPylLeuAsnProValVal 20
Db 535 ATGACTAATAAAAAAGGCGCTGAACCCCTACACCAATGGCTGGAATTAATCCAGTTGTG 594
Oy 21 G1YA1AValTyrSerProAsnPhetTyrAlaGlyThrValLeuLeuGlySerGlnAlaAsnGln 40
Db 595 GGGCGGCGTGTACAGCCCACTTCTATGACAGGACCGTCTGTGTGCGACAGGCCAACAG 654
Oy 41 G1UG1YSerSerMetTyrSerGlyProSerSerLeuValTyrThrSerAlaMetProGly 60
Db 655 GAGGATCTTCCATGATGACGTGGCCCACTTCTATATCTTCTGATGCTTGCCTGGC 714
Oy 61 PheProTyrProAlaAlaAlaThrAlaAlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArg 80
Db 715 TTTCATATCCGGCGCCCACTGCTGACGCTGATACCGAGGGGCTCACCCTCGAGCCCGT 774
Oy 81 G1YA1GThValTyrAsnThrPheAlaGlnAlaAlaAlaProProProProTyrAlaTyr 100
Db 775 GGTCCGACCGTGTACAAACCTTCAAGCTGCGCGCCCGCCCAATCCCGGCTTAT 834
Oy 101 G1YGLYValValTyrGlnGlnProValTyrGlyAsnLysLeuGlnGlyGlyTyrAla 120
Db 835 GCGCGAGTGTGTATACAGCCAGTGTATGCGAATAATGCTACAGGCTGTTACGCT 894
Oy 121 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSerTyrGly 140
Db 895 GCATACCGCTACGCCCGACCCCTGCGCACTGCTGCTACAGTACAGTACAGTACGGA 954
Oy 141 ArgValTyrAlaAlaAlaAspProTyr 148
Db 955 CGAGTTTATGCTGCGACCCCTTAC 978

RESULT 2
; Sequence 1, Application US/09794591
; Patent No. US20010018198A1
; GENERAL INFORMATION:
; APPLICANT: Pulec, Stefan M.
; APPLICANT: Shibata, Hiroki
; TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
; TITLE OF INVENTION: Products Related Thereto and Methods of Using Same
; FILE REFERENCE: CE 3093
; CURRENT APPLICATION NUMBER: US/09/794,591
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/145,391
; PRIOR FILING DATE: 1998-09-01
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2372
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (987)..(1979)
US-09-794-591-1

Alignment Scores:
Pred. No.: 8,48e-46 Length: 2372
Score: 57.00 Matches: 57
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 38.51% Indels: 0
DB: 10 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-794-591-1 (1-2372)
Oy 49 ProSerSerLeuValTyrThrSerAlaMetProGlyPheProTyrProAlaAlaThrAla 68
Db 1716 CCGAGTTCACTTGTATATCTTCTGCAATGCCAGGCTTCCCGATCCAGCAGCACCGCC 1775
Oy 69 AlaAlaAlaTyrArgGlyAlaHisLeuArgGlyArgGlyArgThrValTyrAsnThrPhe 88
```

```
Db 1776 GCGGCGGCTTACCGAGGGCGGACCTGCGAGGCCCGCGGTGACCGTGTACAAACCTTC 1835
Oy 89 ArgAlaAlaAlaProProProProProTyrGlyGlyValValTyr 105
Db 1836 AGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1886

RESULT 3
US-09-878-574-3090
; Sequence 3090, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 3090
; LENGTH: 234
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB3028-016-Q1-B1-B8
US-09-878-574-3090

Alignment Scores:
Pred. No.: 14.6 Length: 234
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 10 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-878-574-3090 (1-234)
Oy 64 ProAlaAlaThrAlaAlaAlaAla 71
Db 22 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 45

RESULT 4
US-09-923-876-1816/c
; Sequence 1816, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Laligudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Ito)
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923,876
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 1816
; LENGTH: 251
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID NO. US20020013958A1 700159573H1
; NAME/KEY: unsure
; LOCATION: 26, 34, 64, 93-94, 151
; OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-1816
```

Alignment Scores:

Pred. No.:	15.6	Length:	251
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	5.41%	Indels:	0
DB:	10	Gaps:	0

US-09-809-545a-2_COPY_1_148 (1-148) x US-09-923-876-1816 (1-251)

QY 89 ArgAlaAlaAlaProProProPro 96

DB 122 CGAGCGCGCGCGCTCCGCGCT 99

RESULT 5

US-09-923-876-4839/C
Sequence 4839, Application US/09923876
Patent No. US20020013958A1
GENERAL INFORMATION:
APPLICANT: Laigudi, Raghunath V.
APPLICANT: Kamigaki, Laura Y. (Ico)
APPLICANT: Sherman, Bradley K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
FILE REFERENCE: PL-0012-1 CON
CURRENT APPLICATION NUMBER: US/09/923,876
CURRENT FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: 09/2298,329
PRIOR FILING DATE: 1999-04-21
PRIOR APPLICATION NUMBER: 60/085,331
PRIOR FILING DATE: 1998-05-05
NUMBER OF SEQ ID NOS: 6332
SOFTWARE: PERL Program
SEQ ID NO: 4839
LENGTH: 263
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20020013958A1 700455882H1
LOCATION: 48, 236, 243, 245
OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-4839

Alignment Scores:			
Pred. No.:	16.3	Length:	263
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	5.41%	Indels:	0
DB:	10	Gaps:	0

US-09-809-545a-2_COPY_1_148 (1-148) x US-09-923-876-4839 (1-263)

QY 64 ProAlaAlaThrAlaAlaAla 71

DB 104 CTTGGGCACTGGCGCTGCAGCG 81

RESULT 6

US-09-983-965-5536/C
Sequence 5536, Application US/09983965
Patent No. US20020137160A1
GENERAL INFORMATION:
APPLICANT: Warren, Weeley C.
APPLICANT: Tao, Nengping
APPLICANT: Byatt, John C.
APPLICANT: Mathiasen, Naeggeppan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 37-21(10297)C
CURRENT APPLICATION NUMBER: US/09/983,965
CURRENT FILING DATE: 2001-10-26

PRIOR APPLICATION NUMBER: US 09/465,231
PRIOR FILING DATE: 1999-12-15
PRIOR APPLICATION NUMBER: US 60/113,678
PRIOR FILING DATE: 1998-12-17
NUMBER OF SEQ ID NOS: 5912
SEQ ID NO: 5536
LENGTH: 355
TYPE: DNA
ORGANISM: Bos taurus
FEATURE:
OTHER INFORMATION: Clone ID: 49-LIB34-024-Q1-E1-E2
US-09-983-965-5536

QY 64 ProAlaAlaThrAlaAlaAla 71

DB 132 CCAGCAGCAACAGCAGCTGCAGCC 109

RESULT 7

US-09-815-343-59
Sequence 59, Application US/09815343
Patent No. US20010055596A1
GENERAL INFORMATION:
APPLICANT: Meagher, Madeleine
APPLICANT: Xu, Jiangchun
APPLICANT: King, Gordon E.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
FILE REFERENCE: 210121.504
CURRENT APPLICATION NUMBER: US/09/815,343
CURRENT FILING DATE: 2001-03-22
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 59
LENGTH: 368
TYPE: DNA
ORGANISM: Homo sapien
US-09-815-343-59

Alignment Scores:			
Pred. No.:	22.2	Length:	368
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	5.41%	Indels:	0
DB:	10	Gaps:	0

US-09-809-545a-2_COPY_1_148 (1-148) x US-09-815-343-59 (1-368)

QY 64 ProAlaAlaThrAlaAlaAla 71

DB 39 CCAGCAGCAACAGCAGCTGCAGCC 62

RESULT 8

US-09-864-761-32302
Sequence 32302, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
GENE EXPRESSION ANALYSIS BY MICROARRAY

```
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 32302
LENGTH: 389
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC024123.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.8
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.7
OTHER INFORMATION: EST HUMAN HIT: BE872908.1, EVALUE 0.00e+00
OTHER INFORMATION: NT HIT: G11436757, EVALUE 1.00e+00
OTHER INFORMATION: SWISSPROT HIT: Q07954, EVALUE 1.00e-23
US-09-864-761-32302

Alignment Scores:
Pred. No.: 23.4 Length: 389
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 10 Gaps: 0

US-09-809-545A-2_copy_1_148 (1-148) x US-09-864-761-32302 (1-389)
QY 64 ProaiaaiaThraiaaiaaiaa 71
Db 324 CCGGGCCGCCACCGCTGCGGCTGCG 347

RESULT 9
US-09-960-352-8373/C
Sequence 8373, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
```

```
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Machialagan, Nageshpan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
FILE REFERENCE: 16511,006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 8373
LENGTH: 407
TYPE: DNA
ORGANISM: Bos taurus
FEATURE:
NAME/KEY: unsure
LOCATION: (359)
OTHER INFORMATION: unsure at all n locations
OTHER INFORMATION: Clone ID: 36-LIB3057-013-Q1-K1-A12
US-09-960-352-8373

Alignment Scores:
Pred. No.: 24.4 Length: 407
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 10 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-960-352-8373 (1-407)
QY 64 ProaiaaiaThraiaaiaaiaa 71
Db 48 CCGGGCCGCCACCGCTGCGGCT 25

RESULT 10
US-10-025-380-659
Sequence 659, Application US/10025380
Publication No. US20020182191A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Lodes, Michael J.
APPLICANT: Secrist, Heather
APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeleine Joy
APPLICANT: Stolk, John A.
APPLICANT: Wang, Tongtong
APPLICANT: Jiang, Yugu
APPLICANT: Smith, Carole L.
APPLICANT: King, Gordon E.
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Skeiky, Yasar A. W.
APPLICANT: Fanger, Gary R.
APPLICANT: Vedvick Thomas S.
APPLICANT: Carter, Darick
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
FILE REFERENCE: 210121,471C14
CURRENT APPLICATION NUMBER: US/10/025,380
CURRENT FILING DATE: 2001-12-19
NUMBER OF SEQ ID NOS: 1129
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 659
LENGTH: 411
TYPE: DNA
ORGANISM: Homo sapiens
US-10-025-380-659

Alignment Scores:
Pred. No.: 24.6 Length: 411
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
```


Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: Gaps: 0

US-09-809-545a-2_COPY_1_148 (1-148) x US-10-025-380-659 (1-411)

QY 64 Proba1a1aThra1a1a1a1a1a1a 71
DB 199 CCAGCAGCAACAGCAGCAGCGCC 222

RESULT 11

US-09-922-217-659
; Sequence 659, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yugui
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922.217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 659
; LENGTH: 411
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-922-217-659

Alignment Scores:
Pred. No.: 24.6 Length: 411
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: Gaps: 0

US-09-809-545a-2_COPY_1_148 (1-148) x US-09-922-217-659 (1-411)

QY 64 Proba1a1aThra1a1a1a1a1a1a 71
DB 199 CCAGCAGCAACAGCAGCAGCGCC 222

RESULT 12

US-09-833-263-659
; Sequence 659, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Stolk, John A.
; APPLICANT: Meagher, Madeleine J.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; FILE REFERENCE: 210121.471C12
; CURRENT APPLICATION NUMBER: US/09/833.263
; CURRENT FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 659
; LENGTH: 411
; TYPE: DNA
; ORGANISM: Homo sapiens

ORGANISM: Homo sapien
US-09-833-263-659

Alignment Scores:

Pred. No.: 24.6 Length: 411
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: Gaps: 0

US-09-809-545a-2_COPY_1_148 (1-148) x US-09-833-263-659 (1-411)

QY 64 Proba1a1aThra1a1a1a1a1a1a 71
DB 199 CCAGCAGCAACAGCAGCAGCGCC 222

RESULT 13

US-09-864-761-11104/c
; Sequence 11104, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864.761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11104
; LENGTH: 464
; TYPE: DNA
; ORGANISM: Homo sapiens

```

; FEATURE:
; OTHER INFORMATION: MAP TO AC002519.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
US-09-864-761-11104

Alignment Scores:
Pred. No.: 27.5 Length: 464
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 10 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-864-761-11104 (1-464)
Oy 64 ProlaAlaThAlaAlaAlaAla 71
Db 228 CCTGCAGCAGCAGCAGCAGCAGCA 205

RESULT 14
US-09-833-790-175
; Sequence 175, Application US/09833790
; Patent No. US20020068288A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongcong
; APPLICANT: Secrist, Heather
; APPLICANT: Monmarch, Raedoh
; APPLICANT: Indriats, Carol Y.
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.512
; CURRENT APPLICATION NUMBER: US/09/833,790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 175
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(501)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-790-175

Alignment Scores:
Pred. No.: 29.5 Length: 501
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 10 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-833-790-175 (1-501)
Oy 90 AlaAlaAlaProProProProPro 97
Db 204 GCTGCTGCTGCTGCTGCTGCTCA 227

RESULT 15
US-09-864-761-15794
; Sequence 15794, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
```

```

; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aomicca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 15794
; LENGTH: 587
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC024123.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.7
US-09-864-761-15794

Alignment Scores:
Pred. No.: 34.2 Length: 587
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.41% Indels: 0
DB: 10 Gaps: 0

US-09-809-545A-2_COPY_1_148 (1-148) x US-09-864-761-15794 (1-587)
Oy 64 ProlaAlaThAlaAlaAlaAla 71
Db 393 CCGGCGGCGCAGCGCTGCGGCGG 416
```

Search completed: March 16, 2003, 05:03:59
Job time : 100 secs

[illegible]

A|Accession: S22254
A|Molecule type: Protein
A|Residues: 2831-2832, 'T', 2834, 'R', 2836-2843, 3002-3014 <DUB>
R|Trueb, J.; Trueb, B.
Biochim. Biophys. Acta 1171, 97-98, 1992
A|Title: The two splice variants of collagen XII share a common 5' end.
A|Reference number: S28811, MUID:93042014, PMID:1120368
A|Accession: S28811
A|Status: Preliminary
A|Molecule type: mRNA
A|Residues: 1-24, 1189-1257, 'S', 1259-1263, 'E', 1265-1280 <TRU>
A|Cross-references: EMBL:X67327
C|Genetics:

A:	Intronns: 2845/3; 2863/3; 2887/3; 2899/3; 2922/1; 2965/1; 3008/1; 3065/1
C:	Superfamily: collagen alpha 1(XII) chain; fibronectin type III repeat homology; von Willebrand factor type A repeat homology <VWA1>
K:	Keywords: alternative splicing; cell binding; coiled coil; connective tissue; disulfide bond
F:	-1-23/Domains: signal sequence #status predicted <SIG>
F:	-24-3124/Product: collagen alpha 1(XII) chain #status predicted <WNT>
F:	-24-1189-3124/Product: collagen alpha 1(XII) chain short splice form #status predicted <IIA>
F:	-24-114/Domains: IIA #status predicted <IIA>
F:	-24-105/Domains: fibronectin type III repeat homology <FN3A>
F:	-137-301/Domains: von Willebrand factor type A repeat homology <VWA1>
F:	-137-425/Domains: IIB #status predicted <IIB>
F:	-137-414/Domains: fibronectin type III repeat homology <FN3B>
F:	-437-601/Domains: von Willebrand factor type A repeat homology <VWA2>
F:	-628-1178/Domains: IIIC #status predicted <IIIC>
F:	-630-711/Domains: fibronectin type III repeat homology <FN3C>
F:	-7721-802/Domains: fibronectin type III repeat homology <FN3D>
F:	-812-895/Domains: fibronectin type III repeat homology <FN3E>
F:	-905-966/Domains: fibronectin type III repeat homology <FN3F>
F:	-995-1076/Domains: fibronectin type III repeat homology <FN3G>
F:	-1086-1169/Domains: fibronectin type III repeat homology <FN3H>
F:	-1197-1361/Domains: von Willebrand factor type A repeat homology <VWA3>
F:	-1184-2295/Domains: IIID #status predicted <IIID>
F:	-1384-1465/Domains: fibronectin type III repeat homology <FN3I>
F:	-1474-1557/Domains: fibronectin type III repeat homology <FN3J>
F:	-1566-1647/Domains: fibronectin type III repeat homology <FN3K>
F:	-1655-1738/Domains: fibronectin type III repeat homology <FN3L>
F:	-1756-1838/Domains: fibronectin type III repeat homology <FN3M>
F:	-1847-1928/Domains: fibronectin type III repeat homology <FN3N>
F:	-1937-2019/Domains: fibronectin type III repeat homology <FN3O>
F:	-2028-2110/Domains: fibronectin type III repeat homology <FN3P>
F:	-2119-2199/Domains: fibronectin type III repeat homology <FN3Q>
F:	-2207-2294/Domains: fibronectin type III repeat homology <FN3R>
F:	-2325-2490/Domains: von Willebrand factor type A repeat homology <VWA4>
F:	-2438-2440/Region: cell adhesion #status predicted
F:	-2459-2500/Domains: IXP, homologous to NC4 domain of type IX collagen #status predicted <IX>
F:	-2571-2902/Domains: collagenous COL2 #status predicted <COL2>
F:	-2859-2801/Region: cell attachment (R-G-D) motif
F:	-2903-2945/Domains: non-collagenous NC2 #status predicted <NC2>
F:	-2946-3048/Domains: collagenous COL1 #status predicted <COL1>
F:	-3049-3124/Domains: non-collagenous NC1 #status predicted <NC1>
F:	-32-1006,1032,1044,1512,1767,2210,2273,2532,2683/Binding site: carbohydrate (Asn) (cova)
F:	-2780,2789,2836,2842,2860,2866,2869,3004,3007/Modified site: hydroxyproline (Pro) #status predicted <MOD>

Alignment Scores:

Pred. No.:	0.589	Length:	3124
Score:	103.00	Matches:	42
Percent Similarity:	34.13%	Conservative:	15
Best Local Similarity:	25.15%	Mismatches:	46
Query Match:	8.85%	Indels:	64
DB:	1	Gaps:	8

```

US-09-809-545A-1 COPY 535 1143 (1-609) x AA0020 (1-3124)

```

QY	184	CCATATCCGCCGCCACTGCTGAGTGCATACCA-----GGGGCTCACCTTCA	234
Db	2094	ProTyr---LysIleThrValAlaValaTYGluAspGlyaspGlyGlnIleunhr	2112
QY	235	GGCCGTGTGCGACCCGG-----	252
Db	2113	GlyenGlyIyrGthrValGlyIeuLeuProProGlnAsnIleTYrIleThrAspGluTrp	2132

Db 154 ThrLeuArgProTyrProSerSerValLeuIleCysSerLeuThrThrProSerThr 173
Qy 194 CGGCACGCTGACAGTGCATACCGAGGGGCTCACCCTTGAGAGCGGTGTGCACCG--- 250
Db 174 AlaProGlyGluGluValIleArgGlyIleArgTyrSerLeuSerIleProSer 193
Qy 251 -----TGACACACCTTCAGAGCTGCGG 274
Db 194 SerAlaProProGluThrAlaAspAspProProGluCysLeuThrProThr----- 210
Qy 275 CGCCCCACCCCAATCCCGGCTATGGCGAGTACGTATCAAGACCAAGTATGCA 334
Db 211 ---ProThrProLeuSerThrProSer-----IleIleArgProGluProGluAla 226
Qy 335 ATAAAT-----TGACACAGGTGGTT-----ACG 358
Db 227 LeuThrLeuProLeuGlySerAlaValSerGluThrThrProThrProIleSerThr 246
Qy 359 CTGCATACCGCTACCGCCAGCCCA---CCCTGCACCTGCTGCTACGTACGACA--- 412
Db 247 ThrThrThrValThrProThrProThrProThrGlyThrGluThrProThrThrPro 266
Qy 413 GTTACGACAGATTATAGTCTGCCGACCCCTTACCAACACACTTGCTCCAGCCCACT 472
Db 267 ILeThrThrThrThrValThrProThrProThrProThrGly---ThrGluThrProThr 285
Qy 473 AGCGCTTGCTGCATGATGCTTTGGCCCTTGACCGATGCCA 517
Db 286 ThrValLeuIleThrThrThrThrThrMetProThrProThrPro 300

RESULT 2

C75607 conserved hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000C:Accession: C75607
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.D.;M.; Shon, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uettersack, T.; Zalewski, C.; Ma
Science 286, 1571-1577, 1999A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: C75607

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-338 <WHI>

A:Cross-references: GB:AE001862; GB:AE001825; NID:g6460468; PIRN:AA12236.1; PID:g646053

A:Experimental source: strain R1

C:Genetics:

A:Gene: DRA0120

A:Map position: 2

C:Superfamily: ymbw protein

US-09-809-545A-1_COPY_535_1143 (1-609) x C75607 (1-338)

Qy 58 GTGGGCGGCTTACAGCCCGACTTATGACAGCAGCGTCTGTTGGCCAGGCCAAC 117

Db 105 LeuGlyLeuGlyValArgAlaProGlyThrAspGlyArgThrAlaLeuAlaLeuArgArgSer 124

Qy 118 CAGGAGGATCTTCATGATCAGTGGCCGACGTTACTTATATCTTCAATGCGCT 177

Db 125 ArgGluAlaLeuAlaAlaAspAspPheProThrGluIleAlaLeuLeuAlaAlaAla 144

Qy 178 GGC---TTTCATATCCGGCCGACCTGTGACGTGACATACCGAGGGGCTCACTTGA 234

Alignment Scores:

Pred. No.:	0.649	Length:	338
Score:	103.00	Matches:	45
Percent Similarity:	37.02%	Conservative:	22
Best Local Similarity:	24.86%	Mismatches:	78
Query Match:	8.85%	Indels:	36
DB:	2	Gaps:	8

Db 145 GlyIuThrProTyrPro-----AlaGlySerValPheAspGlyValLeuAlaGlu 161
Qy 235 GGGCGTGTGCGCACCGTGTACACACCTTACAGCTGGCGGCGCCCAATCCCG 294
Db 162 ProArgGlyValIleValLeu-----ProLeuThr 171
Qy 295 GCTATGCGGAGTATGTATCAAGACCGAGTATGCAATTAATGCTTACAGGGTGT 354
Db 172 IleLeuGlySerSerLeuTyrGlyAlaGluLeuAlaGlyGlu-----LeuGlyValGly 189
Qy 355 TACGCT---GATACCGGTACGCCAGCCCACTGCTGCTGCTCCTACAGTAC 411
Db 190 TyrAlaPheAlaTyrHisPheSerGluIleAspProAlaVal----- 203
Qy 412 AGTTACGACAGATTATGCTGCGGACCCCTACACACACACTTGCTCCAGCCCG--- 468
Db 204 -----AlaValGluThrTyrHisAspHisPheArgProGlyProLeu 217
Qy 469 -----ACCTACGGCGTGTGTCATGATGATGCTTTGGCCCTTGACCGATGCCAAGCT 522
Db 218 AlaGluProTyrTyrAlaIleLeuGlyValAsnAlaLeuAla-----AlaProThr 233
Qy 523 AGAGCCATGCTGATGATGATGAGTGTCTCTTCTTCTTATTCAGGCTAGTATACCA 582
Db 234 GluAspGluAlaArgAspLeuSerLeuThrSerAlaAlaLeuSerLeuGlyIleLeuSer 253
Qy 583 GGG 585
Db 254 Gly 254

RESULT 3

JC4647 K78 protein - rat

N:Alternate names: basic helix-loop-helix protein homolog
C:Species: Rattus norvegicus (Norway rat)

C:Date: 10-May-1996 #sequence_revision 19-Jul-1996 #text_change 20-Jun-2000

C:Accession: JC4647
R:Kume, H.; Matuyama, K.; Tomita, T.; Iwatsubo, T.; Saido, T.C.; Obata, K.

Biochem. Biophys. Res. Commun. 219, 526-530, 1996

A:Title: Molecular cloning of a novel basic helix-loop-helix protein from the rat brain.
A:Reference number: JC4647; MUID:96193685; PMID:8605021

A:Accession: JC4647

A:Molecule type: DNA

A:Residues: 1-381 <KUM>

A:Cross-references: DDBJ:D82868; NID:g1163397; PID:BA11615.1; PID:g1163398

A:Experimental source: brain

C:Comment: This protein is involved in synaptic plasticity, and has a role specific to n

C:Keywords: brain

F:122-134/Region: basic

US-09-809-545A-1_COPY_535_1143 (1-609) x JC4647 (1-381)

Qy 40 TGAATTAATCAAGTGTGTGGCGGCTTACACCGCCGACTTATGACAGCAGCGT 99

Db 171 TrpAlaLeuSerGluIleLeuArgSerGlyAspArgProAspLeuValSerTyrValGlu 190

Qy 100 CTGTGTGCGACGCC----- 114

Db 191 ThrLeuCysGlyLeuSerGluProThrThrAsnLeuValAlaGlyCysLeuGluLeu 210

Qy 115 -----AACGAGAGGATCTTCATGATCAGTGGCCCGCCGCTCACTT 156

Db 211 AsnSerArgAsnPheLeuThrThrGluGluAlaAspGlyValAlaAlaPheThrAlaArg 230

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: March 12, 2003, 19:22:28 ; Search time 23 Seconds
(without alignments)
5090.945 Million cell updates/sec

Title: US-09-809-545a-1_COPY_535_1143
Perfect score: 1164
Sequence: 1 atgactataataaaagccgc.....acaacgccttcgctccatcat 609

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh
-O=/cgn2.1/USPRO.spool/US09809545/runat_11032003_130946_12825/app_query.fasta_1.775
-DB=PIR_73 -QWMT=fastcan -SUFFIX=n2p.rpr -MINMATCH=0.1 -TDOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blowsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US09809545 @CNC 1.1 25 @runat_11032003_130946_12825 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-MARN TIMEOUT=30 -THEADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	9.3	1118	2 A48292	mucin, tracheobron
2	103	8.8	338	2 C75607	conserved hypotet
3	103	8.8	381	2 JC4647	KW8 protein - rat
4	103	8.8	3124	1 A40020	collagen alpha 1(X
5	102.5	8.8	241	2 T22216	hypothetical prote
6	101	8.7	1221	2 T73472	hypothetical prote
7	100.5	8.6	211	2 B89716	protein F4586.3 (1
8	99	8.5	387	2 I50703	transcription fact
9	99	8.5	542	2 T29707	hypothetical prote
10	98.5	8.5	655	1 A55726	RNA-binding protei
11	98.5	8.5	2639	2 T31328	fibroin - Chinese
12	98	8.4	216	2 I51920	mucin - rhesus mac
13	98	8.4	316	2 T31880	hypothetical prote
14	98	8.4	376	2 A26066	segmentation prote

15	97	8.3	811	2 T36581	probable transmem
16	96.5	8.3	139	2 S61885	extensin precursor
17	96.5	8.3	748	2 T04011	hypothetical prote
18	96	8.2	405	4 A61181	homeotic protein H
19	96	8.2	760	2 A45174	eye cell developme
20	95.5	8.2	733	2 T00532	OP protein - Kenne
21	95.5	8.2	1172	1 TSHUP2	chromospondin 2 p
22	95	8.2	418	2 T15142	hypothetical prote
23	94.5	8.1	268	2 B61615	fibroin heavy chai
24	94.5	8.1	469	2 T20047	hypothetical prote
25	94.5	8.1	1006	2 G86292	hypothetical prote
26	94.5	8.1	1964	2 T09059	notch4 - mouse
27	94	8.1	219	2 S42574	adhesive protein -
28	94	8.1	967	2 H86334	T20H2.10 protein -
29	93.5	8.0	294	2 A37232	mucin, tracheal (A
30	93.5	8.0	320	2 S09208	chorion protein s3
31	93.5	8.0	339	2 T25562	hypothetical prote
32	93.5	8.0	839	2 T04859	extensin homolog F
33	93	8.0	197	2 T10081	sperm mitochondria
34	93	8.0	495	1 S31223	transcription fact
35	92.5	7.9	164	2 T15525	hypothetical prote
36	92.5	7.9	348	2 T35248	probable oxidoredu
37	92.5	7.9	656	1 A49358	RNA-binding protei
38	92	7.9	250	2 C87443	conserved hypotet
39	92	7.9	443	2 I49140	p62 ras-GAP associ
40	92	7.9	535	2 S65762	chitinase (EC 3.2.
41	91.5	7.9	325	2 C95952	hypothetical prote
42	91.5	7.9	409	2 A70932	probable ppe prote
43	91.5	7.9	856	2 T13159	EB1-55kDa-associat
44	91.5	7.9	1052	2 T14343	zinc finger RNA bi
45	91.5	7.9	2715	2 T13049	eyelid - fruit fly

ALIGNMENTS

RESULT 1
A48292
mucin, tracheobronchial - dog
C/Species: Canis lupus familiaris (dog)
C/Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 05-Nov-1999
C/Accession: A48292
R/Verma, M.; Davidson, E.A.
Proc. Natl. Acad. Sci. U.S.A. 90, 7144-7148, 1993
A/Title: Molecular cloning and sequencing of a canine tracheobronchial mucin cDNA contai
A/Reference number: A48292; MUID:93348229; PMID:8346228
A/Accession: A48292
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1118 <VER>
A/Cross-references: GB:X69164; NID:9402557; PIDN:CAA48914.1; PID:9402558
A/Note: authors translated the codon CAA for residue 52 as Gly, GAT for residue 94 as Asr
C/Comment: It is uncertain whether Met-1 or Met-52 is the initiator.
C/Keywords: glycoprotein

Alignment Scores:

Pred. No.: 0.228 Length: 1118
Score: 108.00 Matches: 57
Percent Similarity: 40.61% Conservative: 23
Best Local Similarity: 28.93% Mismatches: 76
Query Match: 9.28% Indels: 41
DB: 2 Gaps: 10

US-09-809-545a-1_COPY_535_1143 (1-609) x A48292 (1-1118)

QY 17 CCGGACCCCTACCAAGAGCGTGAATTAATCCAGTTGGCGCGGCTACAGCC 76
DB 117 ProGlnThrProThrProLeuSer-ThrProAlaSerSerSerIlePro-----GlnPr 134
QY 77 CCGACTCTATGACGAGCGGTGCTGTGTCGACGCAACGAGAGGATCTTCAGTGT 136
DB 134 OArg-PheTyrProSerSerValIleuSerAlaValSerGluThrThrProThrThr-Thr 153
QY 137 ACAAGTGCCCGACGATTCATGTAATATATCTTCGCAATGC---CTGGCTTTCATATCCGG 193

THIS PAGE BLANK (USPTO)

```

PN W0200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0654936.
XX 15-SEP-2000; 2000US-0663561.
XX 20-OCT-2000; 2000US-0693325.
XX 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Dermanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX Zhao QH, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI: 2001-476283/51.
XX N-PSDB; AAK53229.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
XX PT useful in diagnosis and gene therapy -
XX
XX PS Claim 20; Page 423; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
XX CC cytokine, cell proliferation or cell differentiation or which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, hematopoiesis regulating
XX CC activity, tissue growth factor activity, immunomodulatory activity and
XX CC activin/inhibin activity and may be useful in the diagnosis and/or
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX CC inflammation.
XX CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
XX CC (AAM80020) are omitted as the relevant pages from the sequence listing
XX CC were missing at the time of publication.
XX
XX SQ Sequence 286 AA;
XX
XX Alignment Scores:
XX Pred. No: 3.86e-16 Length: 286
XX Score: 258.00 Matches: 47
XX Percent Similarity: 96.08% Conservative: 2
XX Best Local Similarity: 92.16% Mismatches: 2
XX Query Match: 22.16% Indels: 0
XX DB: 22 Gaps: 0
XX
XX US-09-809-545A-1_COPY_535_1143 (1-609) x AAM80096 (1-286)
XX
XX 1 ATGACTAATAAAGAGCGCCGTAACCCCTACACCAATGGCTGGAATTAATCCAGTTGG 60
XX |||||||
XX Db 219 MettrhAmhLbVbThrValaenProTyrThraenGlyTrrpLysLeuasnProValVal 238
XX |||||||
XX 61 GGCGGGGTCTACAGCCCGACTTCTATGAGGACGAGTGCTGTGGCCAGGCCAACGAG 120
XX |||||||
XX Db 239 GtAlaValIySerProGluPherIaGlyThrValLeuLeuCyGlnIlaenGln 258
XX |||||||
XX 121 GAGGATCTTCATATACAGTGCGCCCAAGTTCA 153
XX |||||||
XX Db 259 GlugIySerSerMetCtyrSerIaProSerThr 269
XX |||||||

```

Search completed: March 12, 2003, 21:20:40
Job time : 35 secs

QY 514 GCCAAGACTAGAGCCAGCTGATGATGGCTCTGTTCTTTCTTATTCAGGCTAGT 573
Db 340 uPro---LeuAlaMetGluLeuAlaLeuTrp-----ArgVa 351
QY 574 ATATPACCAAGGGGATACACCGTTTG 601
Db 351 ITYrThrGluValAlaThrAlaAspLeu 360
RESULT 14
ABG26323
ID ABG26323 standard; Protein, 388 AA.
AC ABG26323;
XX 18-FEB-2002 (first entry)
DT 18-FEB-2002 (first entry)
DE Novel human diagnostic protein #26314.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
XX MO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
DR N-PSDB; AAS90510.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
PS Claim 20; SEQ ID NO 56682; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 388 AA;
SQ

Alignment Scores: 3.25e-21 Length: 388
Pred. No.:

Score: 311.50 Matches: 90
Percent Similarity: 46.64% Conservative: 14
Best Local Similarity: 40.36% Mismatches: 59
Query Match: 26.76% Indels: 60
DB: 22 Gaps: 10
US-09-809-545A-1_COPY_535_1143 (1-609) x ABG26323 (1-388)
QY 1 ATGACTATATAAAAGCGGTGAACCCCTACACCAATGGCTGGAATTAATTCAGTTGTG 60
Db 186 MetThrAnuLysMetValThrProTyrAlaAsnGlyTrpLysLeuSerProValVal 205
QY 61 GCGCGGCTTACAGCCCGGACTTATGACAGGACGCGTGTGTCGCCAGGCCAAC--- 117
Db 206 GlValAlaValTyrGlyProGluLeuTyrAlaLaseSerPheGlnAlaAspValSerLeu 225
QY 118 ---CAGGAGGATCTTCATGATGACGTGCGCCCGAGTTCATT-----GTATATACTTCT 168
Db 226 GlyLeuAspAlaAlaValProLeuSerGlyArgGlyLeuThrTrpIleProLeu 245
QY 169 GCAATGCGCTGCTTTCATATCCG---GCCGCCACTGCTGACGCTGCATACCGAGGGCT 225
Db 246 IleLeuProGlyPheProTyrProThrAlaAlaThrThrAlaAlaAlaPheArgGlyAla 265
QY 226 CACCTTGAGGCGCGTGTGCGACCGGTGACACACCTTCAGAGCTGGGGCGGCCCACCC 285
Db 266 HisLeuArgGlyArgGlyArgThrValTyrGlyAlaValArg---AlaValProProThr 284
QY 286 CCAATCCCGCGCTATGGCGGAGTA----- 309
Db 285 AlaIleProAlaTyrProGlyValAlaAspLeuProGlyThrAspPheThrValLeuThrSer 304
QY 310 ---GTGTATCAAGACCCAGTGTATGCAATTAATTGCTACAGGGGTGTTACGCTGCATAC 366
Db 305 MetValAspMetGlnProThrAspMetHisSerLeuLeu----- 317
QY 367 CGCTACGCCCGACCCCGCCCTGCTGCTGCT-----ACAGTGACAGT 414
Db 318 ----LeuGlnProGlnPro-ProLeuLeuGlnProLeuGlnProLeuThrValThrVa 335
QY 415 TACGAGCAGATTATGCTGCGACCCCTACACCCACACACTTGCTCCAGCCCGACCTAC 474
Db 335 lMetAlaGlyCysThrGlnProThr----- 345
QY 475 GCGCTTGTCGATGATGATGCTTTGCGCCCTGACCGATGCCAGACTAGAGCCATGCT 534
Db 346 -----MetProLeuProLeuPro---LeuAlaMetGluLe 356
QY 535 GATGATGCGGCTGCTTCTTTCTTATTCAGGCTACTATATACAGGGGATACAC 594
Db 356 uAlaLeuTrp-----ArgValTyrThrGluValAlaThrAl 368
QY 595 CGTTTGG 601
Db 368 aspLeu 370
RESULT 15
AAM80096
ID AAM80096 standard; Protein, 286 AA.
AC AAM80096;
XX 06-NOV-2001 (first entry)
DT Human protein SEQ ID NO 3742.
DE
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
XX Homo sapiens.
OS

Oy	412	TACGGCGTTGGTGCATGAATGCTTTTGGCCCTTGAACCGATGCCAGACTAGAGGCAT	531
Db	24	TyrGIyValGIyAlaMeLAsnAlaPheAlaProLeuThrIspAlaIyVtIrrAysSerHis	43
Oy	532	GCTGATGATGGGCTCGTCTTTCTTCATTCAGGCGGTATATATACCAAGGGGATAC	591
Db	44	AlaAspAlaPValGIyLeuValLeuSerSerLeuGlnAlaSerIleTyrAlGIyGIyTyr	63
Oy	592	AACCGTTTGGCTCCATAT	609
Db	64	AsnArgPheAlaProTyr	69
RESULT 13			
ID	AA156850	standard; Protein; 366 AA.	
XX	AA156850;		
AC	AA156850;		
XX	11-APR-2000	(first entry)	
DT			
XX			
DE		Human RNA binding protein (RNABP)-2 (clone 1250374).	
XX			
KM		RNA binding protein; RNABP; cancer; immune disorder; AIDS; human;	
KW		developmental disorder; acquired immunodeficiency syndrome; RNABP-2;	
KM		inflammation; allergy; diabetes mellitus; seizure disorder.	
XX			
OS		Homo sapiens.	
XX			
PN		US6020164-A.	
XX			
PD		01-FEB-2000.	
XX			
PF		21-OCT-1998; 98US-0176657.	
XX			
PR		21-OCT-1998; 98US-0176657.	
XX			
PA		(INCY-) INCYTE PHARM INC.	
PI		Bandman O, Corley NC, Lu DM, Baughn MR, Tang YT, Guegler KJ;	
XX			
DR		WPI; 2000-146885/13.	
DR		N-PSDB; AA246827.	
PT		Isolated and purified polynucleotide for modulating the expression of	
PT		human RNA binding proteins which play a role in cancer, immune	
XX		disorders and developmental disorders -	
PS		Examples; Columns 45-48; 39pp; English.	
XX			
CC	The present invention provides human RNA binding proteins (RNABP) and		
CC	polynucleotides encoding the proteins. The polypeptides are useful for		
CC	modulating the expression of human RNA binding proteins (RNABP) which		
CC	play a role in cancer, immune disorders and developmental disorders.		
CC	Disorders associated with a decrease of RNABP include: cancers such as		
CC	adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma,		
CC	teratocarcinoma, and, in particular, cancers of the adrenal gland,		
CC	bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia		
CC	gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary,		
CC	pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen,		
CC	testis, thymus, thyroid, and uterus, immune disorders such as acquired		
CC	immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory		
CC	distress syndrome, allergies, ankylosing spondylitis, amyloidosis,		
CC	anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune		
CC	thyroiditis, bronchitis, cholecyelitis, contact dermatitis, Crohn's		
CC	disease, atopic dermatitis, dermatomyositis, diabetes mellitus,		
CC	emphysema, episodic lymphopenia with lymphocytocoxins, erythroblastosis		
CC	fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis,		
CC	hypertension, syndrome, gout, Graves' disease, Hashimoto's thyroiditis,		
CC	hypersplenophilia, irritable bowel syndrome, multiple sclerosis,		
CC	myasthenia gravis, myocardial or pericardial inflammation,		
CC	osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis,		
CC	Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjogren's		

CC	syndrom, systemic anophylaxis, systemic lupus erythematosus, systemic		
CC	sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner		
CC	syndrome, complications of cancer, hemodialysis, and extracorporeal		
CC	circulation, viral, bacterial, fungal, parasitic, protozoal, and		
CC	hematitic infections, trauma, x-linked agammaglobulinemia of Bruton,		
CC	common variable immunodeficiency (CVI), digoxin's syndrome (thymic		
CC	hypoplasia), thymic dysplasia, isolated immunoglobulin (Ig)-A deficiency,		
CC	severe combined immunodeficiency disease (SCID), immunodeficiency with		
CC	thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi		
CC	syndrome, chronic granulomatous diseases, hereditary angioneurotic edema,		
CC	and immunodeficiency associated with Cushing's disease; and developmental		
CC	disorders such as renal tubular acidosis, anemia, Cushing's syndrome,		
CC	achondroplastic dwarfism, Duchenne and Becker muscular dystrophy,		
CC	epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia,		
CC	genitourinary abnormalities, and mental retardation), Smith-Magenis		
CC	syndrome, myelodysplastic syndrome, hereditary mucopolysaccharidosis,		
CC	hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-		
CC	Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus,		
CC	seizure disorders such as Sydenham's chorea and cerebral palsy, spina		
CC	bifida, anencephaly, craniofacial dysmorphism, congenital glaucoma, cataract,		
CC	sensorineural hearing loss, and any disorder associated with cell		
CC	growth and differentiation, embryogenesis, and morphogenesis involving		
CC	any tissue, organ, or system of a subject, e.g., the brain, adrenal		
CC	gland, kidney, skeletal or reproductive system. The present sequence		
CC	represents the RMBP-2 polypeptide.		
CC			
XX			
Sequence	366 AA;		
Alignment Scores:			
Pred. No.:	1,32e-21	Length:	366
Score:	315.50	Matches:	90
Percent Similarity:	49.05%	Conservative:	13
Best Local Similarity:	42.86%	Mismatches:	59
Query Match:	27.10%	Indels:	48
DB:	21	Gaps:	10
US-09-809-545A-1_COPY_535_1143 (1-609) x AAY56850 (1-366)			
OY	1 ATGACTATATAAAGCCGCTGAACCCCTACACCAAGCGCTGGAATTAAATCAAGTTGTG	60	
DB	190 MethranmylsylbMetValThrProTyrAlaAsnGlyTTrpIlybLeuSerProAlaVal	209	
OY	61 GCGCGGCTTACAGCCCCGCACTTCTATGACAGCGACGGTCTGTGTGCCAGGCCAAC--	117	
DB	210 GlyAlaValTyrGlyProGluLeuTyrAlaAlaSerSerPheGlnAlaAspValSerLeu	229	
OY	118 ---CAGAGGAGATCTTCATGATCACTGACGTGGCCCCAGTTCCTT-----GTATATACCTCT	168	
DB	230 GlyAsnAspAlaAlaValProLeuSerGlyTyrArgGlyIlyLeaSnThrTyrIleProLeu	249	
OY	169 GCATGCGCTGGCTTCCATATCGC---GCCGCGACATGCTGCAGTCAATCCAGAGGGCT	225	
DB	250 IleIleProGlyPheProTyrProThrAlaIleThrThrAlaIleAlaIlePheArgGlyAla	269	
OY	226 CACCTTGCAGGCGGTGATCGCACCGCTGTACAACACCTTCAGACTGCGCGCCCCACCC	285	
DB	270 HisLeuArgGlyArgGlyTyrValTyrGlyAlaValArg--AlaValProProThr	288	
OY	286 CCAATCCCGCGCTTATGGCGAGTATAGTCTATCAAGACCCAGTATGCAATMAATTGCTA	345	
DB	289 AlaIleProAlaTyrTyrProGlyValIaPMeTgIn---ProThrAspMetHisSerIleuLeu	307	
OY	346 CAGGTGGTATACGCTGATACCGCTACGGCCAGCCACCCCTGCACATGCTGTGCTCT--	403	
DB	308 -----LeuGlnProGlnPro-ProIlybLeuGlnProIle	318	
OY	404 -----ACAGTGCAGATTACGCGACAGAGTTTATGCTGCGAGCCCTACACACACACA	453	
DB	318 uGlnProLeuThrValThrValMetAlaGlycThrGlnProThrProThr-----	335	
OY	454 CTTGCTCAGAGCCCCACCTTACGGCGGTGTGTCAGTAAATGCTTTTGCCGCTTGACCGAT	513	
DB	336 -----MetProIlybProIle	340	

XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 6810; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AA438642-AA42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, Leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 291 AA:
SQ
Alignment Scores:
Pred. No.: 1,18e-28 Length: 291
Score: 389.00 Matches: 90
Percent Similarity: 57.93% Conservative: 5
Best Local Similarity: 54.88% Mismatches: 38
Query Match: 33.42% Indels: 31
DB: Gaps: 4
US-09-809-545A-1_COPY_535_1143 (1-609) x AA41879 (1-291)
QY 1 ATGACTAATAAAGCGCGTGAACCCCTACACCAATGCTGGAATTAATCCAGTTG 60
DB 147 MetThrAsnLysLysThrGlyAsnProTyrThrAsnGlyTrpLysLeuAsnProValVal 166
QY 61 GCGCGGCTCTACAGCCCGACCTTCTAGCAGGCGGCTCTTGTGCCAGGCCAACCAG 120
DB 167 GAlaAlaValTyrGlyProGluPheTyrAlaValThr----- 178
QY 121 GAGGATCTTCATGTACAGTGCGCCAGTTCATTGTATATCTTCGCAATGCGCTGCG 180
DB 179 -----Gly 179
QY 181 TTTCATATCGGCGCGCAGCTGCTGAGCTGCATACCGAGGGGCTTCACCTTGAGGCCGT 240
DB 180 PheProTyrProThrThrGlyThrAlaValAlaTyrArgGlyAlaHisLeuArgGlyArg 199
QY 241 GGTCCGACCGCTGACAAACCTTCAGAGCTGGCGCGCCGCCCAATCCCGCTAT 300
DB 200 GAlaGAlaValTyrAsnThrPheArgAlaAlaProProProProTyrProThrTyr 219
QY 301 GCGGAGTGTATCAAGAGCCAGTGTGCAATTAATGCTACAGGGGTGTTACGCT 360
DB 220 GAlaAlaValTyrGlnPheGlyPheTyrGlyAlaGluLeuAla---ThrG 238
QY 361 GCATACCGGTACGCCAGCCAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
DB 238 nProThrArgPthrLeuSer---ProLeuGlnArgArgGlnProThrAlaThrValThrAl 257
QY 421 CGAG---TTTATGCTGCCAGCCCTTCACCAACACACTTGTCTCCAGCCCGCTACGGC 477
DB 257 aGluSerThrGlnLeuProThrArgThrIleThrProSerGlyProArgArgProThrAl 277
QY 478 GTTGTCGCA 487
DB 277 aLeuGluPro 280
RESULT 12
ABG29445

ID ABG29445 standard; Protein; 103 AA.
XX
XX AC ABG29445;
XX
XX 18-FEB-2002 (first entry)
XX
XX DE Novel human diagnostic protein #29436.
XX
XX KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX OS Homo sapiens.
XX
XX PN WC200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX
XX PA (HYSE-) HYSEQ INC.
XX
XX PI Drmanac RT, Liu C, Tang YT;
XX
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS93632.
XX
XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX PS Claim 20; SEQ ID NO 59804; 103pp; English.
XX
XX CC The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 103 AA:
XX
XX Alignment Scores:
Pred. No.: 2.72e-24 Length: 103
Score: 342.00 Matches: 64
Percent Similarity: 98.48% Conservative: 1
Best Local Similarity: 96.97% Mismatches: 1
Query Match: 29.38% Indels: 0
DB: Gaps: 0
US-09-809-545A-1_COPY_535_1143 (1-609) x ABG29445 (1-103)
QY 412 AGTTACGCGAGGATTATGCTGCCAGCCCTACACACACTTGTCTCAGCCCGCACCC 471
DB 4 SerTyrGlyArgValTyrAlaAlaAspProTyrThrHisAlaLeuAlaProAlaProThr 23

XX Human; nocotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KM peripheral nervous system; neuropathy; central nervous system; CNS;
 KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KM chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KM leukaemia.
 XX Homo sapiens.
 OS WO200153312-A1.
 PN 26-JUL-2001.
 PD 26-DEC-2000; 2000MO-US34263.
 PF 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0558042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX (HYSE-) HYSEQ INC.
 PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 DR WPI; 2001-442253/47.
 DR N-PSDB; AA161035.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX Example 2; SEQ ID NO 6809; 10078bp; English.
 XX The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AA038642-AA042213) with nocotropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC C.N.S disorders,
 CC assays for receptor activity, arthritis and inflammation, leukaemia and
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 XX Sequence 291 AA:
 SQ
 Alignment Scores:
 Pred. No: 1,18e-28 Length: 291
 Score: 389.00 Matches: 90
 Percent Similarity: 57.93% Conservative: 5
 Best Local Similarity: 54.88% Mismatches: 38
 Query Match: 33.42% Indels: 31
 DB: 22 Gaps: 4
 US-09-809-545a-1_COPY_535_1143 (1-609) x AA041878 (1-291)
 QY 1 ATGACTATAAAAGGCGGTGACCCCTACACCAATGGCGAATGAAATCCAGTTGTG 60
 Db 147 MethionenlyblybtrnglybshpProtyrThrshenglytrpbybLeuabpIvalVal 166
 QY 61 GCGCGGTCTACAGCCCGACTTCTATGACGACGAGTGTGTTGCGAGGCCAACGAC 120

Db 167 GLVALAVALTYRGLYPRGGLUPHEFYRALAVALTHR----- 178
 QY 121 GAGGATCTTCATAGTACAGTGCGCCAGTTCACTGTATATACCTTGCAATGCCCTGC 180
 Db 179 -----Gly 179
 QY 181 TTTCATATCCGCCGCGCACTGTCGACGTGACATACCGAGGGCTACCTTGAGGCCGT 240
 Db 180 PheProTyRProThrThrnglyThnAlaValAlaTyArgGlyAlaHisLeuArgGlyArg 199
 QY 241 GGTGACACCGTGTACACACCTTCAGAGCTGCGCGCCGCCCAATCCGCGCCTAT 300
 Db 200 GLYArgAlaValTYRaaThrPheArgAlaAlaProProProProPoleProThrTyr 219
 QY 301 GCGCGAGTAGTGTATCAAGACCGAGTGTATGCAATTAATGCTACAGGGTGTACGCT 360
 Db 220 GLYAlaValAlaValTYRnglnabpGlyPheTyRGLYAlaGluLeuGlu-Ala---ThrG1 238
 QY 361 GCATACCGCTACGCCCGCCAGCCCGCTGCCACGTGCTGCTACAGTGAACGTTACGA 420
 Db 238 nProThrAspThrLeuSer---ProLeuGlnArgArgGlnProThrAlaThrAl 257
 QY 421 CGAG---TTTATGTGTCGCGACCCCTACACACACACTTGTCTCAGGCCCTACGCG 477
 Db 257 aGluSerThrGlnLeuProThrArgThrIleThrProSerGlyProArgArgProThrAl 277
 QY 478 GTTGTGTCGA 487
 Db 277 aLeuGluPro 280
 RESULT 11
 ID AA041879 standard; Protein; 291 AA.
 XX
 AC AA041879;
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 6810.
 XX
 KM Human; nocotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KM peripheral nervous system; neuropathy; central nervous system; CNS;
 KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KM chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KM leukaemia.
 OS Homo sapiens.
 PN WO200153312-A1.
 PD 26-JUL-2001.
 PF 26-DEC-2000; 2000MO-US34263.
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0558042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX (HYSE-) HYSEQ INC.
 PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 DR WPI; 2001-442253/47.
 DR N-PSDB; AA161035.

RESULT 9
ABG29447

ID ABG29447 standard; Protein; 705 AA.

AC ABG29447;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #294338.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

2000 1999 1998 1997 1996 1995 1994 1993 1992 1991 1990 1989 1988 1987 1986 1985 1984 1983 1982 1981 1980 1979 1978 1977 1976 1975 1974 1973 1972 1971 1970 1969 1968 1967 1966 1965 1964 1963 1962 1961 1960 1959 1958 1957 1956 1955 1954 1953 1952 1951 1950 1949 1948 1947 1946 1945 1944 1943 1942 1941 1940 1939 1938 1937 1936 1935 1934 1933 1932 1931 1930 1929 1928 1927 1926 1925 1924 1923 1922 1921 1920 1919 1918 1917 1916 1915 1914 1913 1912 1911 1910 1909 1908 1907 1906 1905 1904 1903 1902 1901 1900 1899 1898 1897 1896 1895 1894 1893 1892 1891 1890 1889 1888 1887 1886 1885 1884 1883 1882 1881 1880 1879 1878 1877 1876 1875 1874 1873 1872 1871 1870 1869 1868 1867 1866 1865 1864 1863 1862 1861 1860 1859 1858 1857 1856 1855 1854 1853 1852 1851 1850 1849 1848 1847 1846 1845 1844 1843 1842 1841 1840 1839 1838 1837 1836 1835 1834 1833 1832 1831 1830 1829 1828 1827 1826 1825 1824 1823 1822 1821 1820 1819 1818 1817 1816 1815 1814 1813 1812 1811 1810 1809 1808 1807 1806 1805 1804 1803 1802 1801 1800 1799 1798 1797 1796 1795 1794 1793 1792 1791 1790 1789 1788 1787 1786 1785 1784 1783 1782 1781 1780 1779 1778 1777 1776 1775 1774 1773 1772 1771 1770 1769 1768 1767 1766 1765 1764 1763 1762 1761 1760 1759 1758 1757 1756 1755 1754 1753 1752 1751 1750 1749 1748 1747 1746 1745 1744 1743 1742 1741 1740 1739 1738 1737 1736 1735 1734 1733 1732 1731 1730 1729 1728 1727 1726 1725 1724 1723 1722 1721 1720 1719 1718 1717 1716 1715 1714 1713 1712 1711 1710 1709 1708 1707 1706 1705 1704 1703 1702 1701 1700 1699 1698 1697 1696 1695 1694 1693 1692 1691 1690 1689 1688 1687 1686 1685 1684 1683 1682 1681 1680 1679 1678 1677 1676 1675 1674 1673 1672 1671 1670 1669 1668 1667 1666 1665 1664 1663 1662 1661 1660 1659 1658 1657 1656 1655 1654 1653 1652 1651 1650 1649 1648 1647 1646 1645 1644 1643 1642 1641 1640 1639 1638 1637 1636 1635 1634 1633 1632 1631 1630 1629 1628 1627 1626 1625 1624 1623 1622 1621 1620 1619 1618 1617 1616 1615 1614 1613 1612 1611 1610 1609 1608 1607 1606 1605 1604 1603 1602 1601 1600 1599 1598 1597 1596 1595 1594 1593 1592 1591 1590 1589 1588 1587 1586 1585 1584 1583 1582 1581 1580 1579 1578 1577 1576 1575 1574 1573 1572 1571 1570 1569 1568 1567 1566 1565 1564 1563 1562 1561 1560 1559 1558 1557 1556 1555 1554 1553 1552 1551 1550 1549 1548 1547 1546 1545 1544 1543 1542 1541 1540 1539 1538 1537 1536 1535 1534 1533 1532 1531 1530 1529 1528 1527 1526 1525 1524 1523 1522 1521 1520 1519 1518 1517 1516 1515 1514 1513 1512 1511 1510 1509 1508 1507 1506 1505 1504 1503 1502 1501 1500 1499 1498 1497 1496 1495 1494 1493 1492 1491 1490 1489 1488 1487 1486 1485 1484 1483 1482 1481 1480 1479 1478 1477 1476 1475 1474 1473 1472 1471 1470 1469 1468 1467 1466 1465 1464 1463 1462 1461 1460 1459 1458 1457 1456 1455 1454 1453 1452 1451 1450 1449 1448 1447 1446 1445 1444 1443 1442 1441 1440 1439 1438 1437 1436 1435 1434 1433 1432 1431 1430 1429 1428 1427 1426 1425 1424 1423 1422 1421 1420 1419 1418 1417 1416 1415 1414 1413 1412 1411 1410 1409 1408 1407 1406 1405 1404 1403 1402 1401 1400 1399 1398 1397 1396 1395 1394 1393 1392 1391 1390 1389 1388 1387 1386 1385 1384 1383 1382 1381 1380 1379 1378 1377 1376 1375 1374 1373 1372 1371 1370 1369 1368 1367 1366 1365 1364 1363 1362 1361 1360 1359 1358 1357 1356 1355 1354 1353 1352 1351 1350 1349 1348 1347 1346 1345 1344 1343 1342 1341 1340 1339 1338 1337 1336 1335 1334 1333 1332 1331 1330 1329 1328 1327 1326 1325 1324 1323 1322 1321 1320 1319 1318 1317 1316 1315 1314 1313 1312 1311 1310 1309 1308 1307 1306 1305 1304 1303 1302 1301 1300 1299 1298 1297 1296 1295 1294 1293 1292 1291 1290 1289 1288 1287 1286 1285 1284 1283 1282 1281 1280 1279 1278 1277 1276 1275 1274 1273 1272 1271 1270 1269 1268 1267 1266 1265 1264 1263 1262 1261 1260 1259 1258 1257 1256 1255 1254 1253 1252 1251 1250 1249 1248 1247 1246 1245 1244 1243 1242 1241 1240 1239 1238 1237 1236 1235 1234 1233 1232 1231 1230 1229 1228 1227 1226 1225 1224 1223 1222 1221 1220 1219 1218 1217 1216 1215 1214 1213 1212 1211 1210 1209 1208 1207 1206 1205 1204 1203 1202 1201 1200 1199 1198 1197 1196 1195 1194 1193 1192 1191 1190 1189 1188 1187 1186 1185 1184 1183 1182

2 X
4 X

2000

DR N-PSDB; AAS93634.

PT New isolated poly

PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 20; SEQ ID No 59806; 103pp; English

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridization probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques for restoring normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at http://wipo.int/pub/published_pat_sequences.

SQ Sequence 705 AA.

Alignment Scores:

Pred. No.: 4e-30

SCORER:	405.2
PERCENT	37.38
STRENGTH	

Best Local Similarity: 34.43

Query Match: 34.84

DB: 22

US-09-809-545A-1_COPY_535_1143 (1-609) X ABG29447 (1-705)

QY	157	GTAAATATCTTGGAAATAGCTGGCTTTCATATCCGAGCGGCGACCTGCTGACATGAC	216
Db	190	ValGluIleAsnSerValProGlyPheProTyrProAlaIleThrIleAlaIleAlaTyr	209
QY	217	CGAGGAGCTACCTTCGAGGCGCTGGTGGACCCGTTACAAACCTTCAGAGCTGGCGG	276
Db	210	ArgGlyIleAlaIleAsnLeuArgGlyArgGlyArgThrValTyrAsnThrPheArgIleAlaIle	229
QY	277	CCCCCCCCCAATCCCCGCTTATGGCGGAGTA-----	309
Db	230	ProProProProIleProAlaTyrGlyGlyLeuThrPAsnLeuValAspLeuProGluTyr	249
QY	310	-----GTGAT-----	315
Db	250	PheAlaLeuSerAspAspAspPheAsnIleLeuLeuLeuValCysIleArgValTyrThr	269
QY	316	-----CAAGAG-----	321
Db	270	GlyGlnGlnGlnIuThrSerLeuCysHisCysProAlaGlnGlnGlnIAsnLysArgAspLeu	289
QY	322	-----CGAGTGATGGC 333	
Db	290	GlnLeuLysMetPheValAspAlaGlnCysProPheGlnLeuGlyProVal-----	307
QY	334	AATAAATTGGTACAGGGGNGTACGCTGCATACCGCTACCGCCAGCCGACCTGGCCAT	393
Db	308	ProArgLeuGlyGlnGlyGlyTyrAlaAlaTyrArgTyrArgIleGlnProThrProAlaThr	327
QY	394	GCTGCTGCTTACAGTGAC-----	411
Db	328	AlaAlaAlaTyrSerAspArgLeuSerLeuMetTyrGlyAspGlySerGluIleValArg	347
QY	411	-----	411
Db	348	LysGlyArgAsnGlnPheValPheValAlaAlaAspGluIleSerCysAsnThrSerAla	367
QY	411	-----	411
Db	368	GlyCysTyrProSerTyrGlnLeuGluAlaSerSerGlyValGlnGlyValAlaThrSer	387
QY	411	-----	411
Db	388	AlaPheHisPheSerIleProThrSerGlnProThrHisAlaIleAsertyIleLysThr	407
QY	411	-----	411
Db	408	LeuAspSerGlyAspAlaMetAlaAspLysIleAspThrAlaProAlaSerMetGluPhe	427
QY	412	-----AGTTACGAGCAAGTTATGCTGGCGGACCCCTACCAACACACA	453
Db	428	SerPheGlnGlnGlyGlnSerTyrGlyArgValTyrAlaIleAspProTyrHisHisAla	447
QY	454	CTTGCTCCAGCCCCACCTACGAGCGGTGGTGCCATGATAGCTTTTGGCGCCCTTGACCAT	513
Db	448	LeuAlaProAlaProThrTyrGlyValGlyIleMetAsp-----	466
QY	514	GCCAAAGCTAGAGCCATGCTGAATGATGTGGTCTCGTTCTTTCTTATTCAGAGCTAGT	573
Db	461	-----ArgGlnTyr-----SerSerLeuCysValSer	465
QY	574	ATATACCAAGGGGGA 588	
Db	470	ValCysThrIleGly 474	

RESULT 10

AA041878
ID: AA041878 standard. Protein: 291 AA

ADAM1418/8 Standard, 221 mV

AA AC AAM41878:

XX

DE	Human polypeptide SEQ ID NO 6809.
XX	
DI	22-OCI-2001 (A1100-0001)

CC Note: The sequence data for this patent did not form part of the printed
CC specification.

SQ Sequence 268 AA;

Alignment Scores

Pred. No.:	1.51e-42	length:	266
Score:	534.00	Matches:	10
Percent Similarity:	67.68%	Conservative:	6
Best Local Similarity:	64.02%	Mismatches:	23
Query Match:	45.88%	Indels:	30
DB:	22	Gaps:	4

US-09-809-545A-1_COPY_535_2143 (1-609) X AAM40092 (1-268)

QY	1	ATGACTATAA	AAAAAGCCG	TGAACCCCT	CAACACCAAT	GGCTGGAAATTAAT	CAACAGTTGTG	60
Db	134	Met	Thr	Asn	Leu	Ser	Thr	60
QY	61	GGCCGCGCT	CAACAGCC	CCGCACTT	CTATGCAG	GCACAGCGTGCT	GTGTGGCCAG	120
Db	154	Gly	Ala	Val	Tyr	Gly	Pro	120
QY	121	GAGGAGAT	CTTCATG	TACAGT	GGGCCAG	TTCACTGTAT	TACTTTCGA	180
Db	166	-----	-----	-----	-----	-----	Gly	166
QY	181	TTTTCATAT	CCGGCCG	CCACTG	CTGCAG	TTGCATAT	CCGAGGGCT	240
Db	167	Phe	Pro	Tyr	Pro	Thr	Thr	240
QY	241	GGTGGCA	CCCGTGA	CAACACCT	TACAGAG	CTGGGGGCC	CCCAACCCCA	300
Db	187	Gly	Arg	Ala	Val	Tyr	Asn	300
QY	301	GGCGGAG	TAGTAT	CAAGAC	CCAGTGTAT	GGCAATAA	ATTGCTAC	360
Db	207	Gly	Ala	Val	Val	Tyr	Gln	360
QY	361	GCAATCC	GGCTAC	GGCCGCA	ACCCTCG	CACTGTG	TGCTTAC	420
Db	226	Ala	Tyr	Arg	Tyr	Ala	Gln	420
QY	421	CGAGTTAT	---GCT	GGCGAC	CCCTTAC	CCACACAC	ACACTTGC	480
Db	245	Arg	Gly	Tyr	Ala	Ala	Ala	480
QY	478	GTTTGCT	GCCATG	489				
Db	265	Ile	Gly	Thr	Met	268		

RESULT 8
ABG29446
ID ABG29446 standard; Protein; 101 AA

AC ABG29446;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #29437.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens

PN W0200175067-A2.

PD 11-OCT-2001

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167

PA (HYSE-) HYSEQ INC.

PI Dymanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

2000

PT New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -

PS Claim 20; SEQ ID No 59805; 103pp; English

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polynuclease chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AB60010-AB93037 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at <http://wipo.int/pub/published.pct> sequences.

Sequence 101 AA;

Alignment Scores:	
Pred. No.:	2,86e-31
Score:	415.00
Percent Similarity:	83.65%
Best Local Similarity:	80.77%
Query Match:	35,65%
DB:	22
Length:	101
Matches:	84
Conservative:	3
Mismatches:	13
Indels:	
Gaps:	2
	4

US-09-809-545A-1_COPY_535_1143 (1-609) X ABG29446 (1-101

QY	205	GCAGCTGCATACCGAGGGGCTCACTTCGAGAGCGGTGTCGACCGGTGAACAACCTTC	26
Db	1	AlaAlaAlaIatYrAGGlyAlaHisIleuAaGlyAlaGlyAlaThrValIyrAenThrPhe	20
QY	265	AGAGCTCGCGCGCCCCCACCACCCCAATCCCGGCTATGCGGAGATGATGATCAAGAGCA	324
Db	21	ArgAlaAlaAlaIarProIarProIarIarIarIarIatYrGlyValValIarIarIarPoly	40
QY	325	GTGTATGGCAATAAATTCCTACAGAGGTGGTACGTGCATATACCGGTACCGCCAGCCACC	384
Db	41	PheYrYrGly-AlaAspIarIetYrGlyValMet---LeuHisIarIarIarIarProSerLeuP	59
QY	385	CCTGCCACTGCTGCTGCTCTACAGTACGACAGTATACGACAGAGTTTATG---CTGCCAGACCC	444
Db	59	oleuProIeuProuIeuProuIarValThrValThrAspGluPneMetThrPhe-SerPro	79
QY	442	TACACACACACTTGTCTCCAGCCCCCACCCTACGAGGCTTGCGCATGAATGCTTTGCG	501
Db	79	eusEriProuIeuThrCysSerSerProHisIeuAurGArGTrPcYrHisGluCysPneCysT	99
QY	502	CCCTTGAC 509	
Db	99	hrPheasp 101	

PR 03-AUG-2000; 2000US-06531450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 XX
 PA (HYSEQ-) HYSEQ INC.
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AU, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Wehrman R, Drmanac RT;
 XX
 XX MPI; 2001-442253/47.
 DR N-PSDB; AA159249.
 XX
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 XX Example 5; SEQ ID NO 3238; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AA138642-AA142213) with neurotropic,
 CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 XX
 SQ Sequence 237 AA;
 XX
 Alignment Scores:
 Pred. NO.: 1.55e-42 Length: 237
 Score: 534.00 Matches: 105
 Percent Similarity: 67.68% Conservative: 6
 Best Local Similarity: 64.02% Mismatches: 23
 Query Match: 45.88% Indels: 30
 DB: Gaps: 4
 US-09-809-545A-1_COPY_535_1143 (1-609) x AA140093 (1-237)
 QY 1 ATGACTATATAAAGGCGGTGAACCCCTACACCAATGGCTGGAATTAATTCAGTTG 60
 DB 103 MettrAanlyslvYThgIYAsProIYrThrsnslYlTPlYsLeuAnProAlVal 122
 QY 61 GCGCGGCTACAGCGCCGACCTTCATGACGACGGGCTGTGTGCCAGGCACACAG 120
 DB 123 GYAlaValTYrGlyPProGluPheTYrAlaValThr----- 134
 QY 121 GAGGAGCTTCATGATAGACAGTGGGCCCCAGTTCACTGTATATACTTGCAATCCTGCG 180
 DB 125 -----Gly 135
 QY 181 TTTCATATCGGGCGCCACGCTGACGTGACGTATACGAGGGGCTACCTTCGAGCCGT 240
 DB 136 PheProTYrProThrThrGlyThrAlaValAlaTYrGlyAlaHisLeuAArgGlyArg 155
 QY 241 GGTGCACCGGTATACAAACCTTCAGAGCTGGGGGCCCCACCCCAATCCCGAGCTAT 300
 DB 156 GYAlaGAlaValTYrAsnThrPheArgAlaAlaProPProPProPProIleProThrTYr 175
 QY 301 GCGGAGTAGTATCAAGAGCACTGTATGGCAATAATTCAGAGGCTGTATAGCT 360
 DB 176 GYAlaValValTYrGlnAspGlyPheTYrGlyAlaGlu---IleTYrGlyGlyTYrAla 194
 QY 361 GCATACCGCTACGCCACGCCACCCCTGCACCTGCTGCTTCACTAGACAGTACGGA 420

Db 195 Ala1YrAlaGTYrAlaAGlnProAla---AlaAlaAlaAlaAlaIaIaYrSerAspSerTyrGly 213
QY 421 CGAGTTTAT---GCTGCCGACCCCTTACCACACACACTTGTCTCCAGCCCCCACTTACGGC 477
Db 214 ArgAla1YrAlaAlaAlaAspProTyrHisThrIleGlyProAlaAlaIaThrTyrSer 233
QY 478 GTTGTGCCATG 489
Db 234 IleGlyThrMet 237
RESULT 7
ID AAM40092
AA AAM40092 standard; Protein: 268 AA.
XX AAM40092;
XX 22-OCT-2001 (first entry)
DE Human polypeptide SEQ ID NO 3237.
XX Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokineic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
XX Homo sapiens.
OS WO200153312-A1.
PN 26-JUL-2001.
PD
XX 26-DEC-2000; 2000MO-US34263.
PF
XX 21-JAN-2000; 2000US--0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-AUG-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSEQ-) HYSEQ INC.
PA
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Dimaac RT;
XX
DR WPI: 2001-442253/47.
XX N-PSDB: AAI59248.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 5; SEQ ID NO 3237; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM3864-AAM42213) with nootropic,
CC immunosuppressant and cyostatic activity. The polynucleotide are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activation/inhibit activity, chemotactic/chemokineic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.


```

QY 61 GCGCGGTCTACAGCCCGCACTTCTATGACAGGACAGTGTCTTGTGCGAGGCAACAG 120
DB 259 GtAlaValValTyrSerProGluPheThrAlaGlyThrValLeuLeuCyGlnAlaIasnGln 278
QY 121 GAGGATCTTCATGATACAGTGGCCAGTTCACCTTGATATACTTCTGCAATGCTGCGC 180
DB 279 GluGlySerSerMetTyrSerAlaProSerSerLeuValTyrThrSerAlaMetProGly 298
QY 181 TTTCCATATCCGCGCGCCAGCTGCTGACATGATCCAGGGGCTCACCCTTGAGGCGCT 240
DB 299 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyAlaIasnMetArgGlyArg 318
QY 241 GATGCGACCGGTATACACACCTTCAGAGCTGGCGGCGCCCGCCCAATCCCGGCGCTAT 300
DB 319 GtAlaGthrValTyrAsnThrPheArgAlaAlaAlaProProProProIleProAlaTyr 338
QY 301 GCGCGAGTAGTGTATCAAGACGCACTGTATGGCAATAAATTGCTACAGGGGTGTTACGCT 360
DB 339 GtGlyValValTyrGlnAspGlyPheTyrGlyAlaAsp---IleTyrGlyGlyTyrAla 357
QY 361 GCATCCGCTACGCCCGCCAGCCCTGCGACTGCTGCTGCTAGAGTACAGCACTTACAGCA 420
DB 358 AlATyrArgTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSerTyrGly 377
QY 421 CGAGTTATGCTGCGCGACCCCTACACACACACACTTGCTCCAGCCCGCCCACTACGGCGTT 480
DB 378 ArgValTyrAlaAlaAspProTyrHisAlaLeuAlaAlaProAlaProThrTyrGlyVal 397
QY 481 GGTGCATGATGCTTTTGGCGCCCTTGACCGATGCCCAAGACTAGAGAGCCATGCTGATGAT 540
DB 398 GtAlaMetAlaAlaPheAlaPheAlaProLeuThrAspAlaTyrThrArgSerHisAlaAsp 417
QY 541 GGGGCTCGCTCTTCTTCATGAGGCTATATATACCAAGGGGAGACAAACGCTTT 600
DB 418 ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrArgGlyGlyTyrAsnArgPhe 437
QY 601 GCTCCATAT 609
DB 438 AlaProTyr 440

RESULT 4
ID AAY81462 standard; Protein; 330 AA.
XX
AC AAY81462;
XX
DT 03-JUN-2000 (first entry)
XX
DE Human ataxin-2 binding protein (A2BP).
XX
KW Ataxin-2 binding protein; A2BP; human; RNA-binding; cell polarisation;
KW neuronal plasticity; cellular degeneration signal transduction pathway;
KW selective RNA transport; spinocerebellar ataxia type-2;
KW hyperproliferative disorder.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 1023..1312
FT notes="This region contains the regions necessary
FT for binding to ataxin-2"
XX
XX WO200012710-A1.
XX
XX 09-MAR-2000.
XX
XX 01-SEP-1999; 99WO-US20156.
XX
XX 01-SEP-1998; 98US-0145391.
XX
XX (CEDA-) CEDARS SINAI MEDICAL CENT.
XX

```

```

PI Puls SM, Shibata H;
XX
XX WPI: 2000-237873/20.
DR N-PSDB: AAA07075.
XX
XX Nucleic acids encoding an ataxin-2 binding protein useful for
PT inhibiting the expression of active proteins from the SCA2 gene for the
PT treatment of spinocerebellar ataxia type-2 -
XX
XX Claim 19; Page 77-78; 82pp; English.
XX
XX This sequence represents cDNA encoding human ataxin-2 binding protein
XX (A2BP). Nucleotide sequences encoding human A2BP were originally
XX isolated in an adult brain cDNA library using the yeast two hybrid
XX method. The ligand of A2BP, ataxin-2, is a 40 kD protein of unknown
XX function that is encoded by the SCA2 gene located on chromosome 12. SCA2
XX has been linked to the autosomal dominant neurodegenerative disorder
XX spinocerebellar ataxia type-2. Individuals afflicted with the disease
XX exhibit CAG triplet expansion in the SCA2 gene, resulting in ataxin-2
XX containing a polyglutamine stretch of about 35-39 residues, whereas that
XX of normal individuals contains approximately 22 contiguous glutamine
XX residues. A2BP and ataxin-2 are components of a cellular degeneration
XX signal transduction pathway. The pathogenic expanded form of ataxin-2 has
XX a higher affinity for A2BP relative to normal ataxin-2; the presence of
XX the expanded form is likely to promote degeneration. A2BP and ataxin have
XX also been found to have a role in gene regulation. The binding of A2BP to
XX ataxin-2 plays an important role in controlling gene expression via the
XX targeting of transport of specific RNAs, selective RNA transport being
XX mediated via the RNA binding domains of A2BP. A2BP is expressed very
XX early in embryonic development. Both ataxin-2 and A2BP are able to bind
XX CC RNA, and are essential components of the RNA localisation network that
XX establishes cellular polarity in embryogenesis. In highly differentiated,
XX CC polarised cells such as neurons, A2BP and ataxin-2 have a similar
XX CC function and are required for neuronal plasticity. A2BP nucleic acids may
XX be used for the recombinant production of A2BP proteins or fragments
XX thereof according to standard methodologies. For example, an A2BP protein
XX CC with an ataxin-2 or RNA binding capability but no signal transduction
XX CC function can be used as a dominant negative inhibitor of the cellular
XX CC degeneration signal transduction pathway. A2BP proteins with a signal
XX CC transduction function can be used to treat hyperproliferative disorders
XX (e.g., cancer) via stimulation of the cellular degeneration pathway.
XX
XX Sequence 330 AA;
XX
XX Alignment Scores:
XX Pred. No.: 4,28e-51 Length: 330
XX Score: 624.00 Matches: 119
XX Percent Similarity: 89.78% Conservative: 4
XX Best Local Similarity: 86.86% Mismatches: 12
XX Query Match: 53.61% Indels: 2
XX DB: 21 Gaps: 0
XX
XX US-09-809-545A-1_COPY_535_1143 (1-609) x AAY81462 (1-330)
QY 1 ATGACTAATAAAGCCGTGAACCCCTTACACCAATGCTGGAATTAATTCAGTGTG 60
DB 196 MetThrAsnLysLysThrValAsnProTyrThrAsnGlyTyrPlyLeuAsnProValVal 215
QY 61 GCGCGGTCTACAGCCCGCACTTCTATGACAGGACAGTGTCTTGTGCGAGGCAACAG 120
DB 216 GtAlaValValTyrSerProGluPheThrAlaGlyThrValLeuLeuCyGlnAlaIasnGln 235
QY 121 GAGGATCTTCATGATACAGTGGCCAGTTCACCTTGATATACTTCTGCAATGCTGCGC 180
DB 236 GluGlySerSerMetTyrSerAlaProSerSerLeuValTyrThrSerAlaMetProGly 255
QY 181 TTTCCATATCCGCGCGCCAGCTGCTGACATGATCCAGGGGCTCACCCTTGAGGCGCT 240
DB 256 PheProTyrProAlaAlaThrAlaAlaAlaTyrArgGlyAlaIasnMetArgGlyArg 275
QY 241 GGTGCGACCGGTATACACACCTTCAGAGCTGGCGGCGCCCGCCCAATCCCGGCGCTAT 300
DB 276 GtAlaGthrValTyrAsnThrPheArgAlaAlaAlaProProProProIleProAlaTyr 295

```

CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 397 AA;

Alignment Scores:

Pred. No.:	6,58e-88	Length:	397
Score:	1009.50	Matches:	190
Percent Similarity:	95.57%	Conservative:	4
Best Local Similarity:	93.60%	Mismatches:	8
Query Match:	86.73%	Indels:	1
DB:	22	Gaps:	1

US-09-809-545A-1_COPY_535_1143 (1-609) x AAB92573 (1-397)

1 ATGACTATATAAAGGCGGTGAACCCCTACACCAATGCGTGAATTAATCCAGTTGTG 60
196 MetThrAsnIysLysThrValAsnProTyrThrAsnGlyTyrIleAsnProValVal 215
61 GGGCGGTCTACAGCCCGGCACTTCTATGACGACGCGTGTGTCGACGACCAACG 120
216 G1yAlaValTyrSerProGluPheTyrAlaGlyThrValLeuLeuCyG1AlaAsnGln 235
121 GAGGAGTCTTCAGATGATACAGTGGCCCGGCACTTCTATGATTAATCTTCAGATGCTGCGC 180
236 G1uG1ySerSerMetTyrSerAlaProSerSerLeuValTyrThrSerAlaMetProGly 255
181 TTTCATATCCGGCGCGGCACTGCTGACGCTGCATACGAGGAGGCTACCTTCGAGCGCGT 240
256 PheProTyrProAlaAlaThrAlaAlaAlaAlaTyrAlaGlyAlaIleAsnGlyArg 275
241 GGTGCAACCGTGTACACACCTTCAGAGCTGCGGCGGCGGCGGCGGCGGCTAT 300
276 G1yArgThrValTyrAsnThrPheArgAlaAlaAlaProProProAlaProAlaTyr 295
301 GGGCGAGTACGTATCAAGACCGCACTGATGCAATTAATGCTACAGGCTGTTACGCT 360
296 G1yG1yAlaValTyrGlnAspGlyPheTyrG1yAlaAsp---IleTyrG1yG1yTyrAla 314
361 GCATACGCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTAT 420
315 AlaTyrArgTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSerTyrGly 334
421 CGAGTTATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTAT 480
335 ArgAlaTyrAlaAlaAspProTyrGlnIleAlaIleuAlaProAlaProThrTyrGlyAla 354
481 GGTGCGCATGAATGCTTTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTAT 540
355 G1yAlaMetAsnAlaPheAlaProLeuThrAspAlaTyrGlnSerPheAlaAspAsp 374
541 GTGGGTCCGTTCTTTCTTCATGCGAGGCTAGTATATCAAGAGGCGGATACACGTTT 600
375 ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrArgGlyGlyTyrAsnArgPhe 394
601 GCTCCATAT 609
395 AlaProTyr 397

RESULT 3
AAM79112
ID AAM79112 standard; Protein: 440 AA.

XX AAM79112;
AC
XX
XX 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 1774.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
XX MO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001MO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0654936.
XX 15-SEP-2000; 2000US-0663561.
XX 20-OCT-2000; 2000US-0693325.
XX 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Xie AJ, Yang Y, Wehrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
XX
XX N-PSDB; AAK52245.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX
PS Claim 20; Page 4126-4127; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78333-AAK80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX
XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
XX (AAM80020) are omitted as the relevant pages from the sequence listing
XX were missing at the time of publication.

SQ Sequence 440 AA;

Alignment Scores:

Pred. No.:	6.77e-88	Length:	440
Score:	1009.50	Matches:	190
Percent Similarity:	95.57%	Conservative:	4
Best Local Similarity:	93.60%	Mismatches:	8
Query Match:	86.73%	Indels:	1
DB:	22	Gaps:	1

US-09-809-545A-1_COPY_535_1143 (1-609) x AAM79112 (1-440)

1 ATGACTATATAAAGGCGGTGAACCCCTACACCAATGCGTGAATTAATCCAGTTGTG 60
239 MetThrAsnIysLysThrValAsnProTyrThrAsnGlyTyrIleAsnProValVal 258

XX	Alzheimer's disease; gene therapy.
XX	
OS	Rattus norvegicus.
XX	
PN	WO200174901-A2.
XX	
PD	11-OCT-2001.
XX	
PP	23-MAR-2001; 2001WO-US09555.
XX	
PR	31-MAR-2000; 2000US-193548P.
PR	14-MAR-2001; 2001US-0809545.
XX	
PA	(SCIO-) SCIOS INC.
XX	
PI	Stanton LW, White RT;
XX	
DR	WPI; 2002-010779/01.
DR	N-PSDB; AAS94693.
XX	
PT	Novel secreted factor polypeptide useful for treating cardiac diseases
PT	such as arteriosclerosis, myocardial infarction, inflammatory diseases
PT	such as asthma, stroke, and rheumatoid arthritis and renal diseases -
XX	
ES	Claim 18; Fig 1; 189pp; English.

CC The invention relates to rat secreted factor polypeptides and the
CC polynucleotides encoding them. The sequences are useful for treating
CC cardiac, renal or inflammatory diseases. These include cardiac diseases
CC such as congestive heart failure, myocarditis, dilated congestive
CC cardiomyopathy, angina pectoris, myocardial infarction, cardiac
CC arrhythmia, pulmonary hypertension, arteriosclerosis, atherosclerosis and
CC cardiac tumours, renal diseases such as glomerulonephritis, nephrotic
CC syndrome, renal infarction, hereditary nephritis, polycystic kidney
CC disease, chronic renal failure, renal vein thrombosis and medullary
CC sponge kidney and inflammatory diseases such as asthma, rheumatoid
CC arthritis, osteoarthritis, stroke, psoriasis, restenosis, graft versus
CC host reaction, Crohn's disease, ulcerative colitis and Alzheimer's
CC disease. Sequences AAU70146-AAU70178 represent the secreted factor
CC polypeptides of the invention.

SQ Sequence 203 AA;

Alignment scores:	
Pred. No.:	2.67e-95
Score:	1086.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Matchn:	93.30%
DB:	23
Length:	203
Matches:	203
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-09-809-545A-1_COPY_535_1143 (1-609) X LAU70146 (1-203)

QY	1	ATGACTAATAAAAGAGCCGTAACACCCCTTACACCAATGGCTGGAAATTAAATCCAGTTGTG	60
Db	1	MetThrAsnLysLysAlaAlaAsnProTyrTrpHisGlyLysLysLeuAsnProVal	20
QY	61	GGCGCGGTCTACAGGCCCGGCACTTCTATGCAAGGACGCGTGTGTGTGCGAGGCCAACAG	120
Db	21	GlyAlaValTyrSerProAspPheTrpHisGlyTrpValLeuLeuGlyGlnAlaAsnGln	40
QY	121	GAGGGATCTTCATGTACAGTGGCCCCAGTTCACTTGATATTACTTTGCGAATGCCCTGGC	180
Db	41	GlyGlySerSerMetTyrSerIleProSerSerLeuValTyrTrpSerAlaMetProGly	60
QY	181	TTTCCATATCCGGCGCGCACTGCTCAGCTGCATATCCAGGAGGCTCACTTCGAGGCCGT	240
Db	61	PheProTyrProAlaIleHisAlaIleHisAlaIleTyrGlnGlyAlaHisIleLeuArgGlyArg	80
QY	241	GGTGGCAGCGTGTACACACCTTCAGAGCTGGGGGCCCCCAACCCCAATCCGCGCTAT	300
Db	81	GlyArgTrpHisValTyrAsnHisPheMetArgAlaAlaIleProProProIleProIleAlaTyr	100

Qy	301	GGCGAGAGTGTATCAAGAGCCAGGTATAGCGCAATTAATGGCTACAGGGGNGTTAGGCT	360
Db	101	GlyGlyValValTyrGlnGlnProValTyrClyYenbYsLeuLeuGlnGlyGlyTyrAla	120
Qy	361	GCATACCGCTACGGCCAGCCACCCCTGCACCTGCTCTGCTTACACTGACAGTTAGGGA	420
Db	121	AlaTyrAlaGlyTyrAlaGlnProThrProAlaThrAlaAlaAlaTyrSerAspSerTyrGly	140
Qy	421	CGAGTTTATGCTGCGCAGACCCCTACACACACACTTGTCTCCAGGCCCCACCTACGGGCTT	480
Db	141	ArgValTyrAlaAlaAspProTyrHisHisThrLeuAlaProAlaProThrTyrGlyVal	160
Qy	481	GTGCGCAGATGCTTTTGGCCCTTACCGCATGCCATAGACCAATGCTGATGAT	540
Db	161	GlyAlaMetAsnAlaPheAlaProLeuThrAspAlaYsThrArgSerHisAlaAspPhe	180
Qy	541	GTGGCTCTCGTTCTTTCTTCATTGACAGCTAGTATATACCAAGGGGATACCAACCGTTTT	600
Db	181	ValGlyLeuValLeuSerSerLeuGlnAlaSerIleTyrGlnGlyGlyTyrAsnArgPhe	200
Qy	601	GCTCCATAT 609	
Db	201	AlaProTyr 203	

RESULT 2	
AAB92573	
ID	AAB92573 standard; Protein; 397 AA

AC AAB92573;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:10787

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036

PR 11-JAN-2000; 2000JP-0118776

PR 09-JUN-2000; 2000JP-0241899

PA (HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX
4

XX
XX

full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the protein encoded by the

PT full-length cDNAs -

PS Claim 8; SEQ ID 10787; 2537pp + CD ROM; English

CC The present invention describes primer sets for synthesizing 5602
CC ϵ 11 Jacobson codons defined in the specification where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the 3' terminal strand of a polyucleotide which comprises or

the 5602 nucleotide sequences defined in the specification, where oligonucleotide comprises at least 15 nucleotides; or (b) a combination of

CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to,

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 12, 2003, 19:21:58 ; Search time 32 Seconds

(without alignments)
5071.852 Million cell updates/sec

Title: US-09-809-545a-1_COPY_535_1143

Perfect score: 1164

Sequence: 1 atgactacaataaaaaagcgctf.....acaacgcttcttcacat 609

Scoring table:

BLOSUM62			
Xgapop 10.0 ,	Xgapext 0.5		
Ygapop 10.0 ,	Ygapext 0.5		
Fgapop 6.0 ,	Fgapext 7.0		
Delop 6.0 ,	Delext 7.0		

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816540

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ n2p.model -DEV=xjh
-O=/cgn2.1/USFTO.epool/US09809545/runat_11032003_130945_12298/app_query.fasta_1.775
-LOOPEXT=0 -UNITs=b1ts -START=1 -END=-1 -MATRIX=blomsub2 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=80 -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09809545 @CGN 1.53 @runat_11032003_130945_12298 -NCPUs=3
-NO_XLXLY -NO_MMAP -LARGEOUTPUT -NEG_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneeq/geneeqp-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneeq/geneeqp-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneeq/geneeqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1086	93.7	297	23 AAU70146	Rat secreted facto
2	1009.5	86.7	397	22 AAB92573	Human protein sequ
3	1009.5	86.7	440	22 AAM79112	Human protein SEQ
4	624	53.6	310	21 AAY81462	Human ataxin-2 bin
5	534	45.9	135	22 ABB02058	Human human diagno
6	534	45.9	237	22 AAM40093	Human polypeptide
7	534	45.9	268	22 AAM40092	Human polypeptide
8	415	35.7	101	22 ABB29446	Novel human diagno
9	405.5	34.8	705	22 ABB29447	Novel human diagno
10	389	33.4	291	22 AAM41878	Human polypeptide
11	389	33.4	291	22 AAM41879	Human polypeptide
12	342	29.4	103	22 ABB29445	Novel human diagno
13	315.5	27.1	366	21 AAY56850	Human RNA binding
14	311.5	26.8	388	22 ABB26323	Novel human diagno
15	258	22.2	286	22 AAM80096	Human protein SEQ
16	205.5	17.7	151	22 ABB26319	Novel human diagno
17	134	11.5	123	22 AAB06790	Human polypeptide
18	129	11.1	367	22 ABB69047	Drosophila melanog
19	116	10.0	3063	23 ABB90762	Human Tumor Endot
20	116	10.0	3118	22 AAU27790	Human full-length
21	110.5	9.5	219	22 ABB60573	Drosophila melanog
22	110	9.5	474	22 AAU04831	Micromonospora eve
23	108.5	9.3	244	22 AAM38864	Human polypeptide
24	108.5	9.3	297	22 AAM40650	Human polypeptide
25	103.5	8.9	939	22 AAM78860	Human protein SEQ
26	103.5	8.9	943	22 AAM79844	Human protein SEQ
27	101.5	8.7	441	22 AAB93634	Human protein sequ
28	101.5	8.7	950	22 ABB71513	Drosophila melanog
29	101	8.7	172	22 ABB70066	Drosophila melanog
30	100.5	8.6	349	21 AAY85691	Murine SIM-1 prote
31	100.5	8.6	349	22 AAB49332	Human TSP1 domain
32	100.5	8.6	966	23 AAU80189	Drosophila melanog
33	100	8.6	830	22 ABB59880	Mouse ischaemic co
34	98.5	8.5	655	23 ABB57126	Novel human diagno
35	98.5	8.5	1471	22 ABB19528	Novel human diagno
36	98.5	8.5	4561	22 ABB30203	Novel human diagno
37	98.5	8.5	9222	22 ABB21064	Novel human diagno
38	98	8.4	376	22 ABB59035	Drosophila melanog
39	97.5	8.4	828	22 ABB58368	Drosophila melanog
40	97	8.3	179	22 ABB67727	Drosophila melanog
41	97	8.3	250	22 AAM79002	Human protein SEQ
42	97	8.3	301	22 AAM79986	Human protein SEQ
43	96.5	8.3	206	22 ABB29368	Novel human diagno
44	96.5	8.3	716	22 ABB58703	Drosophila melanog
45	96.5	8.3	748	23 ABB92874	Herbicide-activ

ALIGNMENTS

RESULT 1

AAU70146 standard; Protein; 203 AA.

AAU70146;

12-MAR-2002 (first entry)

Rat secreted factor protein encoded by DNA clone P0184_D11.

Rat; secreted factor polypeptide; cardiac disease; renal disease; kidney;
inflammatory disease; congestive heart failure; myocarditis; asthma;
dilated congestive cardiomyopathy; angina pectoris; cardiac arrhythmia;
myocardial infarction; pulmonary hypertension; arteriosclerosis; stroke;
atherosclerosis; cardiac tumour; glomerulonephritis; nephrotic syndrome;
renal infarction; hereditary nephritis; polycystic kidney disease;
chronic renal failure; renal vein thrombosis; medullary sponge kidney;
rheumatoid arthritis; osteoarthritis; psoriasis; restenosis;
graft versus host reaction; Crohn's disease; ulcerative colitis;

THIS PAGE BLANK (USPTO)

```

; FILE REFERENCE: P3430R1C70
; CURRENT APPLICATION NUMBER: US/10/176,482
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 465
; LENGTH: 3582
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-176-482-465

```

```

Query Match 3.0%; Score 18; DB 9; Length 3582;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 276 GCCCCACCCCCCAATCCC 293
Db 3096 GCCCCACCCCCCAATCCC 3113

```

```

RESULT 13
US-10-176-757-465
; Sequence 465, Application US/10176757
; Publication No. US20030022297A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Zhang, Zemin I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C66
; CURRENT APPLICATION NUMBER: US/10/176,757
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 465
; LENGTH: 3582
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-176-757-465

```

```

Query Match 3.0%; Score 18; DB 9; Length 3582;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 276 GCCCCACCCCCCAATCCC 293
Db 3096 GCCCCACCCCCCAATCCC 3113

```

```

RESULT 14
US-10-176-913-465
; Sequence 465, Application US/10176913
; Publication No. US20030022298A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
US-10-176-913-465

```

```

; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C66
; CURRENT APPLICATION NUMBER: US/10/176,913
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 465
; LENGTH: 3582
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-176-913-465

```

```

Query Match 3.0%; Score 18; DB 9; Length 3582;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 276 GCCCCACCCCCCAATCCC 293
Db 3096 GCCCCACCCCCCAATCCC 3113

```

```

RESULT 15
US-10-180-552-465
; Sequence 465, Application US/10180552
; Publication No. US20030022300A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C153
; CURRENT APPLICATION NUMBER: US/10/180,552
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 465
; LENGTH: 3582
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-180-552-465

```

```

Query Match 3.0%; Score 18; DB 9; Length 3582;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 276 GCCCCACCCCCCAATCCC 293
Db 3096 GCCCCACCCCCCAATCCC 3113

```

```

Search completed: March 16, 2003, 02:35:20
Job time : 80.3694 secs

```

LENGTH: 3582
TYPE: DNA
ORGANISM: Homo Sapien
US-10-175-737-465

Query Match
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 GCCCCACCCCAATCCC 293
DB 3096 GCCCCACCCCAATCCC 3113

RESULT 9
US-10-173-706-465
Sequence 465, Application US/10173706
Publication No. US20030022293A1

GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C7
CURRENT FILING DATE: 2002-06-17
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 465
LENGTH: 3582
TYPE: DNA
ORGANISM: Homo Sapien
US-10-173-706-465

Query Match
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 GCCCCACCCCAATCCC 293
DB 3096 GCCCCACCCCAATCCC 3113

RESULT 10
US-10-175-738-465
Sequence 465, Application US/10175738
Publication No. US20030022294A1

GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C45
CURRENT FILING DATE: 2002-06-19

Prior application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 465
LENGTH: 3582
TYPE: DNA
ORGANISM: Homo Sapien
US-10-175-738-465

Query Match
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 GCCCCACCCCAATCCC 293
DB 3096 GCCCCACCCCAATCCC 3113

RESULT 11
US-10-175-752-465
Sequence 465, Application US/10175752
Publication No. US20030022295A1

GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C60
CURRENT FILING DATE: 2002-06-19
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 465
LENGTH: 3582
TYPE: DNA
ORGANISM: Homo Sapien
US-10-175-752-465

Query Match
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 GCCCCACCCCAATCCC 293
DB 3096 GCCCCACCCCAATCCC 3113

RESULT 12
US-10-176-482-465
Sequence 465, Application US/10176482
Publication No. US20030022296A1

GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

ORGANISM: Homosapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: n,a,t,g or c
US-09-969-708-535

Query Match 3.0%; Score 18; DB 10; Length 332;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 454 CTTGCTCCAGCCCCACC 471
DB 142 CTGCTCCAGCCCCACC 125

RESULT 5
US-10-102-558-1/c
Sequence 1, Application US/10102558
Publication No. US2003008706X1
GENERAL INFORMATION:
APPLICANT: Dai, Ken-Shwo
TITLE OF INVENTION: HUMAN PEN1B-RELATED GENE VARIANT ASSOCIATED WITH LUNG CANCERS
FILE REFERENCE: U 013922-3
CURRENT FILING DATE: 2002-03-20
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 2173
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (128)..(574)
OTHER INFORMATION:
US-10-102-558-1

Query Match 3.0%; Score 18; DB 9; Length 2173;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 271 GCGGCGCCCCCACC 288
DB 1260 GCGGCGCCCCCACC 1243

RESULT 6
US-10-174-590-465
Sequence 465, Application US/10174590
Publication No. US2003008352A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Deenoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C42
CURRENT FILING DATE: 2002-06-18
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 465
LENGTH: 3582
TYPE: DNA
ORGANISM: Homo Sapien

US-10-174-590-465

Query Match 3.0%; Score 18; DB 9; Length 3582;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 276 GCGCCCAACCCCAATCCC 293
DB 3096 GCGCCCAACCCCAATCCC 3113

RESULT 7
US-10-176-758-465
Sequence 465, Application US/10176758
Publication No. US2003008353A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Deenoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C104
CURRENT FILING DATE: 2002-06-21
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 465
LENGTH: 3582
TYPE: DNA
ORGANISM: Homo Sapien
US-10-176-758-465

Query Match 3.0%; Score 18; DB 9; Length 3582;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 276 GCGCCCAACCCCAATCCC 293
DB 3096 GCGCCCAACCCCAATCCC 3113

RESULT 8
US-10-175-737-465
Sequence 465, Application US/10175737
Publication No. US2003001315A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Deenoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C50
CURRENT FILING DATE: 2002-06-19
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 465

Qy	301	GGCGAGTAGTATCAAGACCAAGTATAGGAATAAAATGGTACAGGGTGGTAACT	360
Db	835	GCGGAGTAGTATCAAGACCAAGTATAGGAATAAAATGGTACAGGGTGGTAACT	894
Qy	361	GCATACCGCTACGCCACGCCACCCCTGCCACTGCTGCTGCCTACAGTACAGTTACGA	420
Db	895	GCATACCGCTACGCCACGCCACCCCTGCCACTGCTGCTGCCTACAGTACAGTTACGA	954
Qy	421	CGAGTTATGCTGCCGACCCCTACACACACACTTGCTCAGGCCCCCACTTACGGCGTT	480
Db	955	CGAGTTATGCTGCCGACCCCTACACACACACTTGCTCAGGCCCCCACTTACGGCGTT	1014
Qy	481	GCTGCCAGTAATGCTTTTGGGCCCTTACCCGATGCCAAGTACAGAGCCATCTGATAT	540
Db	1015	GCTGCCAGTAATGCTTTTGGGCCCTTACCCGATGCCAAGTACAGAGCCATCTGATAT	1074
Qy	541	GTGGGTGCGTCTTTCTTCTATGACAGGTAGTATATACCAAGGGGATACACCGTTTT	600
Db	1075	GTGGGTGCGTCTTTCTTCTATGACAGGTAGTATATACCAAGGGGATACACCGTTTT	1133
Qy	601	GCTCCATAT	609
Db	1135	GCTCCATAT	1143

```

RESULT 2
US-09-794-591-1
: Sequence 1, Application US/09794591
: Patent No. US20010018198A1
: GENERAL INFORMATION:
: APPLICANT: Pulet, Stefan M.
: TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
: TITLE OF INVENTION: Products Related Thereto and Methods of Using Same
: FILE REFERENCE: CE 3093
: CURRENT APPLICATION NUMBER: US/09/794,591
: CURRENT FILING DATE: 2001-02-26
: PRIOR APPLICATION NUMBER: 09/145,391
: PRIOR FILING DATE: 1998-09-01
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 2372
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (987)..(1979)
US-09-794-591-1

```

Query Match	11.5%	Score 70;	DB 10;	Length 2372;
Best Local Similarity	100.0%	Pred. No. 6e-28;		
Matches	70;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0
QY	511	GATGCCAAGCTAGAGCCATGCTGATGATGGAGGCTGTTCTTTCTTATGCAAGCT	570	
Db	2133	GATGCCAAGCTAGAGCCATGCTGATGATGGAGGCTGTTCTTTCTTATGCAAGCT	2192	
QY	571	AGTATATACC	580	
Db	2193	AGTATATACC	2202	

RESULT 3
 US-09-783-590-7497
 : Sequence 7497, Application US/09783590
 : Patent No. US20020110850A1
 : GENERAL INFORMATION:
 : APPLICANT: Dillon, Patrick J.
 : APPLICANT: Haselcaine, William A.
 : APPLICANT: Li, Heodong
 : APPLICANT: Rosen, Craig A.

? APPLICANT: Ruben, Steven M.
 ? TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
 ? FILE REFERENCE: PO-16.2C1
 ? CURRENT APPLICATION NUMBER: US/09/783,590
 ? CURRENT FILING DATE: 2000-02-15
 ? PRIOR APPLICATION NUMBER: 08/420,856
 ? PRIOR FILING DATE: 1995-04-12
 ? PRIOR APPLICATION NUMBER: 08/346,731
 ? PRIOR FILING DATE: 1994-11-21
 ? NUMBER OF SEQ ID NOS: 12485
 ? SOFTWARE: PatentIn Ver. 2.0
 ? SEQ ID NO 7497
 ? LENGTH: 378

```

? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc feature
? LOCATION: (31)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (137)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (161)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (165)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (172)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (252)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (266)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (271)
? OTHER INFORMATION: n equals a,t,g, or c
US-09-7497

```

```

Query Match          3.0%; Score 18; DB 10; Length 278;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0;
Gaps 0;
QY 276 GCGCCCAACCCCAATCCC 293
    |||||
Db 115 GCGCCCAACCCCAATCCC 132

```

```

RESULT 4
US-09-969-708-535/C
; Sequence 535, Application US/09969708
; Patent No. US20020102532A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-70
; CURRENT APPLICATION NUMBER: US/09/969,708
; CURRENT FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: US/60/237,606
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,608
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,425
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 658
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 535
; LENGTH: 332
; TYPE: DNA

```

GenCore version 5.1.4_p5_4570
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 15, 2003, 16:12:39 ; Search time 54.6694 Seconds
(without alignments)
7861.696 Million cell updates/sec

Title: US-09-809-545a-1_COPY_535_1143
Perfect score: 609
Sequence: 1 atgacctataaagaagccgt.....acaacgcttcgctccatcat 609

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 501302 seqs, 350932545 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1002604

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- Published Applications NA:*
- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
 - 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
 - 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
 - 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
 - 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
 - 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
 - 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
 - 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
 - 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
 - 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
 - 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
 - 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
 - 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
 - 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	609	100.0	1340	10 US-09-809-545a-1	Sequence 1, Appl1
2	70	11.5	2372	10 US-09-794-591-1	Sequence 1, Appl1
3	18	3.0	278	10 US-09-783-590-7497	Sequence 7497, Ap
4	18	3.0	332	10 US-09-969-708-535	Sequence 535, App
5	18	3.0	2173	9 US-10-102-558-1	Sequence 1, Appl1
6	18	3.0	3582	9 US-10-174-580-465	Sequence 465, App
7	18	3.0	3582	9 US-10-176-758-465	Sequence 465, App
8	18	3.0	3582	9 US-10-175-737-465	Sequence 465, App
9	18	3.0	3582	9 US-10-173-706-465	Sequence 465, App
10	18	3.0	3582	9 US-10-175-738-465	Sequence 465, App
11	18	3.0	3582	9 US-10-175-752-465	Sequence 465, App
12	18	3.0	3582	9 US-10-176-482-465	Sequence 465, App
13	18	3.0	3582	9 US-10-176-757-465	Sequence 465, App
14	18	3.0	3582	9 US-10-176-913-465	Sequence 465, App
15	18	3.0	3582	9 US-10-180-552-465	Sequence 465, App
16	18	3.0	3582	9 US-10-180-557-465	Sequence 465, App
17	18	3.0	3582	9 US-10-173-700-465	Sequence 465, App
18	18	3.0	3582	9 US-10-174-572-465	Sequence 465, App
19	18	3.0	3582	9 US-10-174-579-465	Sequence 465, App

20	18	3.0	3582	9 US-10-174-582-465	Sequence 465, App
21	18	3.0	3582	9 US-10-174-588-465	Sequence 465, App
22	18	3.0	3582	9 US-10-175-739-465	Sequence 465, App
23	18	3.0	3582	9 US-10-175-740-465	Sequence 465, App
24	18	3.0	3582	9 US-10-175-743-465	Sequence 465, App
25	18	3.0	3582	9 US-10-176-488-465	Sequence 465, App
26	18	3.0	3582	9 US-10-176-492-465	Sequence 465, App
27	18	3.0	3582	9 US-10-176-747-465	Sequence 465, App
28	18	3.0	3582	9 US-10-176-750-465	Sequence 465, App
29	18	3.0	3582	9 US-10-176-985-465	Sequence 465, App
30	18	3.0	3582	9 US-10-176-987-465	Sequence 465, App
31	18	3.0	3582	9 US-10-176-991-465	Sequence 465, App
32	18	3.0	3582	9 US-10-176-993-465	Sequence 465, App
33	18	3.0	3582	9 US-10-176-993-465	Sequence 465, App
34	18	3.0	3582	9 US-10-184-658-465	Sequence 465, App
35	18	3.0	3582	9 US-10-173-695-465	Sequence 465, App
36	18	3.0	3582	9 US-10-173-697-465	Sequence 465, App
37	18	3.0	3582	9 US-10-173-705-465	Sequence 465, App
38	18	3.0	3582	9 US-10-174-576-465	Sequence 465, App
39	18	3.0	3582	9 US-10-174-585-465	Sequence 465, App
40	18	3.0	3582	9 US-10-174-586-465	Sequence 465, App
41	18	3.0	3582	9 US-10-175-747-465	Sequence 465, App
42	18	3.0	3582	9 US-10-176-481-465	Sequence 465, App
43	18	3.0	3582	9 US-10-176-485-465	Sequence 465, App
44	18	3.0	3582	9 US-10-176-487-465	Sequence 465, App
45	18	3.0	3582	9 US-10-176-493-465	Sequence 465, App

ALIGNMENTS

RESULT 1
US-09-809-545a-1
Sequence 1, Application US/09809545A
Patent No. US20020110804A1
GENERAL INFORMATION:
APPLICANT: Stanton, Lawrence W.
APPLICANT: White, R. Tyler
TITLE OF INVENTION: SECRETED FACTORS
FILE REFERENCE: SCIOS.017A
CURRENT APPLICATION NUMBER: US/09/809,545A
CURRENT FILING DATE: 2001-03-14
NUMBER OF SEQ ID NOS: 84
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1340
TYPE: DNA
ORGANISM: Rattus norvegicus
US-09-809-545a-1

Query Match	100.0%	Score 609;	DB 10;	Length 1340;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 609;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGACTAATATAAAGCGCGTGACCCCTACACCAATGGCTGGAATTAATCCAGTTGTG	60	
DB	535	ATGACTAATATAAAGCGCGTGACCCCTACACCAATGGCTGGAATTAATCCAGTTGTG	594	
QY	61	GGCCCGCTACACGCCCGAGTTCTATGACGACGCGTGTGCGACGACCAACAG	120	
DB	595	GGCCCGCTACACGCCCGAGTTCTATGACGACGCGTGTGCGACGACCAACAG	654	
QY	121	GAGGATCTTCATGATGACAGTGGCCCAAGTTCATCTTATATATCTTGCAATCCCTGGC	180	
DB	655	GAGGATCTTCATGATGACAGTGGCCCAAGTTCATCTTATATATCTTGCAATCCCTGGC	714	
QY	181	TTTCCATATCCGCGCCGCTACCTGTCAGCTGCATATCCGAGGGCTCACTTCGAGCGCT	240	
DB	715	TTTCCATATCCGCGCCGCTACCTGTCAGCTGCATATCCGAGGGCTCACTTCGAGCGCT	774	
QY	241	GCTGCACCGCTATCAACACCTTACAGAGCTGCGGCGCCCAACCCCAATCCGCGCTAT	300	
DB	775	GCTGCACCGCTATCAACACCTTACAGAGCTGCGGCGCCCAACCCCAATCCGCGCTAT	834	

THIS PAGE BLANK (USPTO)

STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/578.615A
FILING DATE: 11-JAN-1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 852,852
FILING DATE: 16-MAR-1992
APPLICATION NUMBER: 08/089,996
FILING DATE: 09-JUL-1993
APPLICATION NUMBER: 08/199,779
FILING DATE: 22-FEB-1994
ATTORNEY/AGENT INFORMATION:
NAME: Paul K. Legaard
REGISTRATION NUMBER: 38,534
REFERENCE/DOCKET NUMBER: ISIS-1568
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 1812 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-578-615A-104

Query Match 2.8%; Score 17; DB 3; Length 1812;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 277 CCCCCCCCCCAATCCC 293
DB 328 CCCCCCCCCCAATCCC 312

RESULT 14
PCT-US94-07770-104/C
Sequence 104, Application PC/TUS9407770
GENERAL INFORMATION:
APPLICANT: Nicholas Dean, C. Frank Bennett and
APPLICANT: Russell T. Boggs
TITLE OF INVENTION: Oligonucleotide Modulation of
NUMBER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB
MEDIUM TYPE: STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/07770
FILING DATE: herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 852,852

FILING DATE: March 16, 1992
APPLICATION NUMBER: 08/089,996
FILING DATE: July 9, 1993
APPLICATION NUMBER: 08/199,779
FILING DATE: February 22, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Rebecca Ralph Gaumont
REGISTRATION NUMBER: 35,152
REFERENCE/DOCKET NUMBER: ISIS-1546
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 1812 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US94-07770-104

Query Match 2.8%; Score 17; DB 5; Length 1812;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 277 CCCCCCCCCCAATCCC 293
DB 328 CCCCCCCCCCAATCCC 312

RESULT 15
US-09-146-053-6
Sequence 6, Application US/09146053A
Patent No. 6399349
GENERAL INFORMATION:
APPLICANT: Ryan, James W.
APPLICANT: Sprinkle, Terry Joe Curtis
APPLICANT: Venema, Richard C.
TITLE OF INVENTION: Human Aminopeptidase P Gene
FILE REFERENCE: MCG103
CURRENT APPLICATION NUMBER: US/09/146,053A
CURRENT FILING DATE: 1998-09-02
EARLIER APPLICATION NUMBER: 60/057,854
EARLIER FILING DATE: 1997-09-02
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 45546
TYPE: DNA
ORGANISM: Homo sapiens
US-09-146-053-6

Query Match 2.8%; Score 17; DB 4; Length 45546;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 365 ACCGCTAGCCCAAGCC 381
DB 22991 ACCGCTAGCCCAAGCC 23007

Search completed: March 16, 2003, 02:32:35
Job time : 69.3715 secs

TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 105:
SEQUENCE CHARACTERISTICS:
LENGTH: 1757 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-578-615A-105

Query Match 2.8%; Score 17; DB 3; Length 1757;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 277 CCCCCCCCCCAATCCC 293
Db 273 CCCCCCCCCCAATCCC 257

RESULT 11

PCT-US94-07770-105/C
Sequence 105, Application PC/TUS9407770
GENERAL INFORMATION:
APPLICANT: Nicholas Dean, C. Frank Bennett and
Applicant: Russell T. Boggs
TITLE OF INVENTION: Oligonucleotide Modulation of
Kinase C
NUMBER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz
ADDRESSEE: Mackiewicz & Norris
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB
MEDIUM TYPE: STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/07770
FILING DATE: herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 852,852
FILING DATE: March 16, 1992
APPLICATION NUMBER: 08/089,996
FILING DATE: July 9, 1993
APPLICATION NUMBER: 08/189,779
FILING DATE: February 22, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Rebecca Ralph Gaumond
REGISTRATION NUMBER: 35,152
REFERENCE/DOCKET NUMBER: ISIS-1546
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3439
TELEFAX: (215) 568-3100
INFORMATION FOR SEQ ID NO: 105:
SEQUENCE CHARACTERISTICS:
LENGTH: 1757 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US94-07770-105

Query Match 2.8%; Score 17; DB 5; Length 1757;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 277 CCCCCCCCCCAATCCC 293
Db 273 CCCCCCCCCCAATCCC 257

RESULT 12

US-08-199-779-1/C
Sequence 1, Application US/08199779
Patent No. 5681747
GENERAL INFORMATION:
APPLICANT: Nicholas Dean, C. Frank Bennett
TITLE OF INVENTION: Oligonucleotide Modulation of
Kinase C
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz
ADDRESSEE: Mackiewicz & No. 5681747ris
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB
MEDIUM TYPE: STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/199,779
FILING DATE: herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 089,996
FILING DATE: July 9, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Rebecca Ralph Gaumond
REGISTRATION NUMBER: 35,152
REFERENCE/DOCKET NUMBER: ISIS-1242
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1812 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-199-779-1

Query Match 2.8%; Score 17; DB 1; Length 1812;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 277 CCCCCCCCCCAATCCC 293
Db 328 CCCCCCCCCCAATCCC 312

RESULT 13

US-08-578-615A-104/C
Sequence 104, Application US/08578615A
Patent No. 6015892
GENERAL INFORMATION:
APPLICANT: Nicholas Dean, C. Frank Bennett and Russell, T. Boggs
TITLE OF INVENTION: Oligonucleotide Modulation of Protein Kinase C
NUMBER OF SEQUENCES: 122
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6015892ris LLP
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia

APPLICANT: Rhee, Sang-Ki
APPLICANT: Lee, Eun-Hae
TITLE OF INVENTION: Glucobacter Suboxydans Sorbitol Dehydrogenase, Genes
TITLE OF INVENTION: and Methods of Use Thereof
FILE REFERENCE: 1533 0870000
CURRENT APPLICATION NUMBER: US/09/296,284A
CURRENT FILING DATE: 1999-04-22
NUMBER OF SEQ ID NOS: 87
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 7
LENGTH: 4830
TYPE: DNA
ORGANISM: Glucobacter suboxydans
US-09-296-284-7

Query Match 3.0%; Score 18; DB 4; Length 4830;
Best Local Similarity 100.0%; Pred. No. 6,7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 274 GCGCCCCCAGCCCAATC 291
DB 97 GCGCCCCCAGCCCAATC 114

RESULT 8
US-09-337-307A-10/C
Sequence 10, Application US/09337307A
Patent No. 6432692
GENERAL INFORMATION:
APPLICANT: Bradfield, Christopher A.
APPLICANT: Carver, Lucy A.
APPLICANT: Dunham, Elizabeth E.
TITLE OF INVENTION: Sensitive Biossay For Detecting Agonists Of The Aryl Hydrocarbon
FILE REFERENCE: MARF0105
CURRENT APPLICATION NUMBER: US/09/337,307A
CURRENT FILING DATE: 1999-06-21
NUMBER OF SEQ ID NOS: 19
SOFTWARE: PatentIn version 3.1
SEQ ID NO: 10
LENGTH: 36
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic construct
US-09-337-307A-10

Query Match 2.8%; Score 17; DB 4; Length 36;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 CCACTGCTGACAGTGA 213
DB 30 CCACTGCTGACAGTGA 14

RESULT 9
US-08-199-779-2/C
Sequence 2, Application US/08199779
Patent No. 5681747
GENERAL INFORMATION:
APPLICANT: Nicholas Dean, C. Frank Bennett
TITLE OF INVENTION: Oligonucleotide Modulation of
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS: Kinase C
ADDRESSEE: Woodcock Washburn Kurtz
ADDRESSEE: Mackiewicz & No. 5681747x1s
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB
MEDIUM TYPE: STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/199,779
FILING DATE: herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 089,996
FILING DATE: July 9, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Rebecca Ralph Gaumond
REGISTRATION NUMBER: 35,152
REFERENCE/DOCKET NUMBER: ISIS-1242
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1757 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-199-779-2

Query Match 2.8%; Score 17; DB 1; Length 1757;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 277 CCCCCCCCCCAATCCC 293
DB 273 CCCCCCCCCCAATCCC 257

RESULT 10
US-08-578-615A-105/C
Sequence 105, Application US/08578615A
Patent No. 6015892
GENERAL INFORMATION:
APPLICANT: Nicholas Dean, C. Frank Bennett and Russell, T. Boggs
TITLE OF INVENTION: Oligonucleotide Modulation of Protein KinaseC
NUMBER OF SEQUENCES: 122
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6015892x1s LLP
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/578,615A
FILING DATE: 11-JAN-1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 852,852
FILING DATE: 16-MAR-1992
APPLICATION NUMBER: 08/089,996
FILING DATE: 09-JUL-1993
APPLICATION NUMBER: 08/199,779
FILING DATE: 22-FEB-1994
ATTORNEY/AGENT INFORMATION:
NAME: Paul K. Leggaard
REGISTRATION NUMBER: 38,534
REFERENCE/DOCKET NUMBER: ISIS-1568
TELECOMMUNICATION INFORMATION:

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 29
LENGTH: 615
TYPE: DNA
ORGANISM: Gluconobacter suboxydans
US-09-296-284-29

Query Match 3.0%; Score 18; DB 4; Length 615;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 274 GCGCCCCCACCCTCAATC 291
|||||
Db 48 GCGCCCCCACCCTCAATC 65

RESULT 3
US-09-296-284-28
Sequence 28, Application US/09296284A
Patent No. 6204040
GENERAL INFORMATION:

APPLICANT: Choi, Eun-Sung
APPLICANT: Rhee, Sang-Ki
APPLICANT: Lee, Eun-Hae
TITLE OF INVENTION: Gluconobacter Suboxydans Sorbitol Dehydrogenase, Genes
FILE REFERENCE: 1533.0870000
CURRENT APPLICATION NUMBER: US/09/296,284A
CURRENT FILING DATE: 1999-04-22
NUMBER OF SEQ ID NOS: 87

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 28
LENGTH: 664
TYPE: DNA
ORGANISM: Gluconobacter suboxydans
US-09-296-284-28

Query Match 3.0%; Score 18; DB 4; Length 664;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 274 GCGCCCCCACCCTCAATC 291
|||||
Db 97 GCGCCCCCACCCTCAATC 114

RESULT 4
US-09-930-181-3/c
Sequence 3, Application US/09930181
Patent No. 6455292
GENERAL INFORMATION:

APPLICANT: Origene Technologies
TITLE OF INVENTION: Full-Length Serine Protein Kinase in Brain and Pancreas
FILE REFERENCE: 16U 101 V1
CURRENT APPLICATION NUMBER: US/09/930,181
CURRENT FILING DATE: 2001-08-16
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Version 3.0
SEQ ID NO 3
LENGTH: 3364
TYPE: DNA

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (482) .. (2239)
US-09-930-181-3

Query Match 3.0%; Score 18; DB 4; Length 3364;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 GCGCGCCCCCACCCTCA 288
|||||

Db 3039 GCGGCGCCCCCACCCTCA 3022

RESULT 5
US-09-632-098-1
Sequence 1, Application US/09632098
Patent No. 6420154
GENERAL INFORMATION:

APPLICANT: Sheppard, Paul O.
APPLICANT: Baidur, Nand
APPLICANT: Bishop, Paul D.
TITLE OF INVENTION: MAMMALIAN ADHESION PROTEASE PEPTIDES
FILE REFERENCE: 99-39
CURRENT APPLICATION NUMBER: US/09/632,098
CURRENT FILING DATE: 2000-08-02
NUMBER OF SEQ ID NOS: 26

SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 3431
TYPE: DNA

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (37) ... (2442)
US-09-632-098-1

Query Match 3.0%; Score 18; DB 4; Length 3431;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 GCGCCCCCACCCTCAATCC 293
|||||
Db 2978 GCGCCCCCACCCTCAATCC 2995

RESULT 6
US-09-632-098-3
Sequence 3, Application US/09632098
Patent No. 6420154
GENERAL INFORMATION:

APPLICANT: Sheppard, Paul O.
APPLICANT: Baidur, Nand
APPLICANT: Bishop, Paul D.
TITLE OF INVENTION: MAMMALIAN ADHESION PROTEASE PEPTIDES
FILE REFERENCE: 99-39
CURRENT APPLICATION NUMBER: US/09/632,098
CURRENT FILING DATE: 2000-08-02
NUMBER OF SEQ ID NOS: 26
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 3468
TYPE: DNA

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (37) ... (2472)
US-09-632-098-3

Query Match 3.0%; Score 18; DB 4; Length 3468;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 GCGCCCCCACCCTCAATCC 293
|||||
Db 3015 GCGCCCCCACCCTCAATCC 3032

RESULT 7
US-09-296-284-7
Sequence 7, Application US/09296284A
Patent No. 6204040
GENERAL INFORMATION:
APPLICANT: Choi, Eun-Sung

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 16:12:15 ; Search time 34.3715 Seconds
(without alignments)
5433.755 Million cell updates/sec

Title: US-09-809-545A-1_COPY_535_1143

Perfect score: 609
Sequence: 1 atgactataataaaagcgcgt.....acaacgcttcgctccatcat 609

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	70	11.5	2372	4 US-09-145-391-1	Sequence 29, Appli
2	18	3.0	615	4 US-09-296-284-29	Sequence 28, Appli
3	18	3.0	664	4 US-09-296-284-28	Sequence 1, Appli
4	18	3.0	3364	4 US-09-930-181-3	Sequence 3, Appli
5	18	3.0	3431	4 US-09-632-098-1	Sequence 3, Appli
6	18	3.0	3468	4 US-09-632-098-3	Sequence 7, Appli
7	18	3.0	4830	4 US-09-296-284-7	Sequence 10, Appli
8	17	2.8	36	4 US-09-337-307A-10	Sequence 2, Appli
9	17	2.8	1757	1 US-08-199-779-2	Sequence 105, App
10	17	2.8	1757	1 US-08-578-615A-105	Sequence 105, App
11	17	2.8	1757	5 PCT-US94-07770-105	Sequence 1, Appli
12	17	2.8	1812	1 US-08-199-779-1	Sequence 104, App
13	17	2.8	1812	3 US-08-578-615A-104	Sequence 104, App
14	17	2.8	1812	3 PCT-US94-07770-104	Sequence 104, App
15	17	2.8	45546	4 US-09-146-053-6	Sequence 6, Appli
16	16	2.6	437	1 US-08-631-200-3	Sequence 4, Appli
17	16	2.6	437	1 US-08-631-200-4	Sequence 3, Appli
18	16	2.6	437	1 US-08-828-553-3	Sequence 3, Appli
19	16	2.6	437	1 US-08-828-553-4	Sequence 3, Appli
20	16	2.6	437	1 US-08-828-553-5	Sequence 3, Appli
21	16	2.6	437	2 US-08-922-267A-3	Sequence 3, Appli
22	16	2.6	437	2 US-08-922-267A-4	Sequence 3, Appli
23	16	2.6	437	2 US-08-922-267A-5	Sequence 3, Appli
24	16	2.6	437	2 US-08-936-707A-3	Sequence 3, Appli
25	16	2.6	437	2 US-08-936-707A-4	Sequence 3, Appli
26	16	2.6	437	2 US-08-936-707A-5	Sequence 3, Appli
27	16	2.6	437	2 US-08-936-707A-6	Sequence 3, Appli

28	16	2.6	437	2 US-08-936-706A-3	Sequence 3, Appli
29	16	2.6	437	2 US-08-936-706A-4	Sequence 4, Appli
30	16	2.6	437	2 US-08-936-706A-5	Sequence 5, Appli
31	16	2.6	437	3 US-09-248-203-3	Sequence 3, Appli
32	16	2.6	437	3 US-09-248-203-4	Sequence 4, Appli
33	16	2.6	437	3 US-09-248-203-5	Sequence 5, Appli
34	16	2.6	437	4 US-09-406-071-3	Sequence 3, Appli
35	16	2.6	437	4 US-09-406-071-4	Sequence 4, Appli
36	16	2.6	437	4 US-09-406-071-5	Sequence 5, Appli
37	16	2.6	437	4 US-08-905-223-86	Sequence 86, Appli
38	16	2.6	480	1 US-08-630-592-5	Sequence 5, Appli
39	16	2.6	480	1 US-08-714-991-5	Sequence 5, Appli
40	16	2.6	480	2 US-08-837-029-1	Sequence 5, Appli
41	16	2.6	833	2 US-08-883-526-2	Sequence 1, Appli
42	16	2.6	1366	3 US-08-883-526-2	Sequence 2, Appli
43	16	2.6	1458	4 US-09-291-023A-1	Sequence 1, Appli
44	16	2.6	1605	2 US-09-000-846-1	Sequence 1, Appli
45	16	2.6	1799	4 US-09-810-347-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-145-391-1
; Sequence 1, Application US/09145391
; Patent No. 6194171
; GENERAL INFORMATION:
; APPLICANT: Pulist, Stefan M.
; APPLICANT: Shibata, Hiroki
; TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
; FILE REFERENCE: CE 3093
; CURRENT APPLICATION NUMBER: US/09/145,391
; CURRENT FILING DATE: 1998-09-01
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2372
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (987)..(1979)
US-09-145-391-1

Query Match 11.5%; Score 70; DB 4; Length 2372;

Best Local Similarity 100.0%; Pred. No. 6e-27; Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 511 GATCCCAAGACTAGAGCATGCTGATGATGTGGCTCTGTTCTTTCTTCATTCAGAGCT 570
Db 2133 GATCCCAAGACTAGAGCATGCTGATGATGTGGCTCTGTTCTTTCTTCATTCAGAGCT 2192

Qy 571 AGTATATACC 580
Db 2193 AGTATATACC 2202

RESULT 2
US-09-296-284-29
; Sequence 29, Application US/09296284A
; Patent No. 6204040
; GENERAL INFORMATION:
; APPLICANT: Choi, Eun-Sung
; APPLICANT: Rhee, Sang-Ki
; APPLICANT: Lee, Eun-Hae
; TITLE OF INVENTION: Glucocorticoid Suboxysol Sorbitol Dehydrogenase, Genes
; FILE REFERENCE: 1533.0870000
; CURRENT APPLICATION NUMBER: US/09/296,284A
; CURRENT FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 87

Qy 571 AGTATATACC 580
|||||
Db 158 AGTATATACC 167

RESULT 12
D60520/c 363 bp mRNA linear EST 28-AUG-1995
LOCUS D60520
DEFINITION HM114E06A clone tech human fetal brain polyA+ mRNA (#6535) Homo
sapiens CDNA clone GEN-114E06 3', mRNA sequence.
ACCESSION D60520
VERSION D60520.1 GI:962159
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 363)
AUTHORS Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M.,
Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H., Takachi,
A., Takeda,S., Matanabe,T., Takahashi,E., Hirai,Y., Maekawa,H.,
Shin,S. and Nakamura,Y.
Fujiwara et al. (1995)
Unpublished (1995)
COMMENT Contact: Tsutomu Fujiwara
Otsuka GEN Research Institute
Otsuka Pharmaceutical Co., Ltd
463-10 Kagasuno Kawasuchi-cho, Tokushima, Tokushima, 771-01 Japan
Tel: 0886-65-2888
Fax: 0886-37-1035.

FEATURES
source
1..363
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GEN-114E06"
/clone_1id="Clontech human fetal brain polyA+ mRNA (#6535)"
/note="Male adult, hematopoietic tissue, stem cell"

BASE COUNT 90 a 84 c 97 g 87 t
ORIGIN

Query Match 11.5%; Score 70; DB 14; Length 363;
Best Local Similarity 100.0%; Pred. No. 1.5e-25;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 511 GATGCCAAGACTAGAGCGCATGCTGATGCGGTCTGTTCTTCTTCATTCGAGGCT 570
|||||
Db 201 GATGCCAAGACTAGAGCGCATGCTGATGCGGTCTGTTCTTCTTCATTCGAGGCT 142
|||||

Qy 571 AGTATATACC 580
|||||
Db 141 AGTATATACC 132

RESULT 13
A1095813 367 bp mRNA linear EST 05-OCT-1998
LOCUS A1095813
DEFINITION qb20g11.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
IMAGE:1696868 3', mRNA sequence.
ACCESSION A1095813
VERSION A1095813.1 GI:3434789
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 367)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

FEATURES
source
1..367
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1696868"
/clone_1id="Soares_pregnant_uterus_NbHPU"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pTTT3-Pac; Site 1: Not I;
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
ACCTGAGAGATTGCGCGCCCTTTTCTTTTCTTTTCTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pTTT3 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

BASE COUNT 65 a 116 c 96 g 90 t
ORIGIN

Query Match 11.5%; Score 70; DB 9; Length 367;
Best Local Similarity 100.0%; Pred. No. 1.5e-25;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 511 GATGCCAAGACTAGAGCGCATGCTGATGCGGTCTGTTCTTCTTCATTCGAGGCT 570
|||||
Db 288 GATGCCAAGACTAGAGCGCATGCTGATGCGGTCTGTTCTTCTTCATTCGAGGCT 347
|||||

Qy 571 AGTATATACC 580
|||||
Db 348 AGTATATACC 357

RESULT 14
A1189407 383 bp mRNA linear EST 28-OCT-1998
LOCUS A1189407
DEFINITION qd01n05.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
IMAGE:1172489 3', mRNA sequence.
ACCESSION A1189407
VERSION A1189407.1 GI:3740616
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 383)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert length: 486 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 381.
Location/Qualifiers
1..383
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1722489"
/clone_1id="Soares_pregnant_uterus_NbHPU"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pTTT3-Pac; Site 1: Not I;
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -

Seq primer: M13/PUC Reverse.
 FEATURES
 source
 1. .408
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="UTSW_S3855"
 /clone_lib="UTSW Adult Mouse Skeletal Muscle Library"
 /sex="Pooled"
 /tissue_type="Diaphragm/Hind limb muscles"
 /cell_type="Skeletal muscle"
 /dev_stage="2 months"
 /lab_host="DH5a"
 /note="Vector: PAMPI0 (Gibco); Cloned unidirectionally. Primer: Oligo dt. RNA Isolation: cytoplasmic RNA preps (Mannias); Cloning Technique: CNA Cloning (Clontech), Life Technologies; Average insert size: 1.8 Kb; Insertion site: TACGTCCATGAATTCGAGC--->. Other information regarding entire library may be found at http://pga.smed.edu/Data/Libraries/microarray_cdna_libraries.htm."

BASE COUNT 108 a 104 c 91 g 101 t 4 others
 ORIGIN

Query Match 13.1%; Score 80; DB 12; Length 403;
 Best Local Similarity 100.0%; Pred. No. 8.9e-31;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 AACCTACCAAGATGCGAATTAATCAATGTCGGCGGCTACAGCCCGAC 81
 Db 232 AACCTACCAAGATGCGAATTAATCAATGTCGGCGGCTACAGCCCGAC 291

Qy 82 TTCTATGACGACGCTCT 101
 Db 292 TTCTATGACGACGCTCT 311

RESULT 10 228 bp mRNA linear EST 14-DEC-1993
 LOCUS 228900/c
 DEFINITION HSB87F031 STRATAGENE Human skeletal muscle cDNA library, cat.
 #938215. Homo sapiens cDNA clone B7F03, mRNA sequence.
 ACCESSION 228900
 VERSION 228900.1 GI:434570
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 228)
 Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Devignes, M. D., Duprat, S., Houllgate, R., Jumeau, M. N., Lamy, B., Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pietu, G., Pouliot, Y., Sebastiani-Kabakchis, C. and Tessier, A.
 IMAGE: molecular integration of the analysis of the human genome and its expression
 C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
 9527534
 JOURNAL
 MEDLINE
 COMMENT
 Contact: Genethon
 Genexpress-Genethon
 Genethon Centre de recherche sur le Genome Humain
 1, rue de l'Internationale, BP60 91002 Evry Cedex, FRANCE
 Tel: 33169472800
 Fax: 33160778698
 Email: genexpress@genethon.fr
 strand(-), single read.
 Location/Qualifiers
 1. .228
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="B7F03"
 /clone_lib="STRATAGENE Human skeletal muscle cDNA library, cat. #938215."
 /sex="Female"

/tissue_type="skeletal muscle"
 /dev_stage="19 years"
 /note="Organ: leg muscle; Vector: Lambda ZAPII; Tissue from female, 19 years old, normal leg muscle. Cloning vector is Lambda ZAPII, in vivo excision from Lambda ZAPII to Bluescript SK(+). Genexpress library reference is B. "
 BASE COUNT 64 a 51 c 48 g 64 t 1 others
 ORIGIN

Query Match 11.5%; Score 70; DB 14; Length 228;
 Best Local Similarity 100.0%; Pred. No. 1.3e-25;
 Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 511 GATGCCAAGACTAGAGGACCATGATGATGAGTCTGTTCTTTTCATTGACGCT 570
 Db 201 GATGCCAAGACTAGAGGACCATGATGATGAGTCTGTTCTTTTCATTGACGCT 142

Qy 571 AGTATATACC 580
 Db 141 AGTATATACC 132

RESULT 11 327 bp mRNA linear EST 21-APR-2001
 LOCUS BG186689
 DEFINITION RST5664 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
 ACCESSION BG186689
 VERSION BG186689.1 GI:13708376
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 327)
 Harrington, J. J., Sherf, B., Rundlett, S., Jackson, P. D., Perry, R., Cain, S., Leventhal, C., Thornton, M., Ramchandran, R., Whittington, J., Lerner, U., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith, E., Veloso, N., Kijka, A., Hess, J., Cothren, K., Lo, K., Offendackher, J., Danzig, J. and Ducar, M.
 Creation of genome-wide protein expression libraries using random activation of gene expression
 Nat. Biotechnol. 19 (5), 440-445 (2001)
 21227151
 JOURNAL
 MEDLINE
 COMMENT
 Contact: Scott J. Cain
 Athersys, Inc.
 3201 Carnegie Ave, Cleveland, OH 44115, USA
 Tel: 216 431 9900
 Fax: 216 361 9596
 Email: scai@athersys.com
 High quality sequence stop. 327.
 Location/Qualifiers
 1. .327
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Athersys RAGE Library"
 /cell_line="HT1080"
 /note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

FEATURES
 source
 1. .327
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Athersys RAGE Library"
 /cell_line="HT1080"
 /note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT 99 a 78 c 69 g 81 t
 ORIGIN

Query Match 11.5%; Score 70; DB 12; Length 327;
 Best Local Similarity 100.0%; Pred. No. 1.5e-25;
 Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 511 GATGCCAAGACTAGAGGACCATGATGATGAGTCTGTTCTTTTCATTGACGCT 570
 Db 98 GATGCCAAGACTAGAGGACCATGATGATGAGTCTGTTCTTTTCATTGACGCT 157

REFERENCE 1 (bases 1 to 538)
AUTHORS Ronaldo M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704447
COMMENT On Feb 17, 1998 this sequence version replaced gi:2888390.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@iuii.wisc.wiowa.edu

The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult brain library. cDNA library preparation: M. Fatima Ronaldo, Ph.D. Clone distribution: clones will be available through Research Genetics. This clone is also available through the I.M.A.G.E. Consortium at LNL (info@image.lnl.gov). IMAGE ID=1767889 The following repetitive elements were found in this cDNA sequence: 105-154, >(CAA)n#Simple_repeat
Seq primer: M13 Forward
POLY-A=NO.

FEATURES
SOURCE 1. Location/Qualifiers
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-A0-ar-b-11-0-UI"
/clone_lib="UI-R-A0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; This library consists of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."

BASE COUNT 146 a 115 c 102 g 175 t
ORIGIN

Query Match 15.1%; Score 92; DB 9; Length 538;
Best Local Similarity 100.0%; Pred. No. 4.9e-37;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 489 GAATGCTTTGGCCCTTGACGAGCCAGACTAGGAGCCATGCTGATGGGCTCT 548
DB 284 GAATGCTTTGGCCCTTGACGAGCCAGACTAGGAGCCATGCTGATGGGCTCT 225

CY 549 CGTCTTTCTTCATTGCAGCGTAGTATATACC 580
DB 224 CCGTCTTTCTTCATTGCAGCGTAGTATATACC 193

RESULT 8
LOCUS B1988957 600 bp mRNA linear EST 20-DEC-2001
DEFINITION 4023-45 Mouse E14.5 retina lambda ZAP II library Mus musculus cDNA, mRNA sequence.
ACCESSION B1988957
VERSION B1988957.1 GI:17953947
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 600)

AUTHORS Mu, X., Zhao, S., Pershad, R., Heieh, T.-F., Searpa, A., Wang, S.W., White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.
TITLE Gene expression in the developing mouse retina by EST sequencing and microarray analysis
JOURNAL Nucleic Acids Res. 29 (24), 4983-4993 (2001)
MEDLINE 21671825
COMMENT Contact: Klein WH
Department of Biochemistry and Molecular Biology
University of Texas M.D. Anderson Cancer Center
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
Tel: 713 792 3646
Fax: 713 790 0329.

FEATURES
SOURCE 1. Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="Mouse E14.5 retina lambda ZAP II library"
/tissue_type="neural retina"
/dev_stage="embryonic day 14.5 post-fertilization"
/note="Vector: pAMP10 (Gibco); Cloned unidirectionally. Primer: Oligo dT. RNA isolation: cytoplasmic RNA preps (Mammalian); Cloning Technique: cDNA Cloning (Clontekmp, Life Technologies); Average insert size: 1.8 Kb; Insertion site: TAGCTCCATCGATTCTGAGTGC---. Other information regarding entire library may be found at http://pga.swmed.edu/Data/libraries/microarray_cdna_library.htm."

BASE COUNT 171 a 124 c 122 g 183 t
ORIGIN

Query Match 15.1%; Score 92; DB 13; Length 600;
Best Local Similarity 100.0%; Pred. No. 5e-37;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 489 GAATGCTTTGGCCCTTGACGAGCCAGACTAGGAGCCATGCTGATGGGCTCT 548
DB 450 GAATGCTTTGGCCCTTGACGAGCCAGACTAGGAGCCATGCTGATGGGCTCT 509

CY 549 CGTCTTTCTTCATTGCAGCGTAGTATATACC 580
DB 510 CCGTCTTTCTTCATTGCAGCGTAGTATATACC 541

RESULT 9
LOCUS BG795922 408 bp mRNA linear EST 16-MAY-2001
DEFINITION UTSW_SM385 UTSW Adult Mouse Skeletal Muscle Library Mus musculus cDNA clone UTSW_SM385, mRNA sequence.
ACCESSION BG795922
VERSION BG795922.1 GI:14131492
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 408)
AUTHORS Gallardo, T.D., Schageman, J.J., Pertsemidis, A., Garner, H.R., Williams, R.S. and Shohet, R.V.
TITLE UT Southwestern Medical Center, Adult Mouse Skeletal Muscle cDNA Library
JOURNAL Unpublished (2001)
COMMENT Contact: Schageman JJ
Shohet/Garner Labs
University of Texas Southwestern Medical Center
6000 Harry Hines Blvd., NA2.226, Dallas, TX 75390, USA
Tel: 214 648 1674
Email: Jeff.Schageman@UTSouthwestern.edu
cDNA library constructed by UTSW and the Reynolds Heart Disease Genomic Applications (PGA) and the Reynolds Heart Disease Prevention grants for use in cDNA microarray experiments. Sequence Quality: Sequence ends were trimmed based on percentage of ambiguous base calls or 'N's in windowed segments. Sequencing: First-pass sequencing; ABI Prism 377 sequencer and analysis software.

updated accordingly when that means is determined. The following repetitive elements were found in this cDNA sequence: 132-168, >(CAAA)n\$imple repeat
Seq primer: M13 Forward
POLYA=yes

FEATURES

Source

Location/Qualifiers

```
1. 361
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="UT-M-AM1-a-fv-h-01-0-UT"
/dev_stage="27-32 days"
/lab_host="DHI0B (Life Technologies)"
/notes="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not 1; Site 2: Eco RI; The
NIH_BMAP_MAM_N library is a normalized library constructed
from mouse amygdala. The tag is a string of 5 nucleotides
present between the Not I site and the oligo-dT track.
The library was constructed as described by Bonaldo,
Lennon and Soares, Genome Research 6: 791-806, 1996.
Tissue provided by Ms. Annie Novakovich, Zivic-Miller
Laboratories.
TAG_LIB=NIH_BMAP_MAM_N
TAG_TISSUE=amygdala
TAG_SEQ=GTGAG"
```

BASE COUNT 79 a 69 c 88 g 125 t
ORIGIN

Query Match 25.0%; Score 152; DB 9; Length 361;
Best Local Similarity 100.0%; Pred. No. 1.3e-68;
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 429 TCCTGCCGACCCCTACACACACTTCTCCAGCCGCCCTACGCGGTGTGTCAT 488
DB 358 TCCTGCCGACCCCTACACACACTTCTCCAGCCGCCCTACGCGGTGTGTCAT 299
OY 489 GAATGCTTTGGCCCTTACCGATGCCAAGACTAGAGCCATGTGATGTGGGTCT 548
DB 299 GAATGCTTTGGCCCTTACCGATGCCAAGACTAGAGCCATGTGATGTGGGTCT 239
OY 549 CGTCTCTTCTTCAATGCGAGGCTAGTATATACC 580
DB 238 CGTCTCTTCTTCAATGCGAGGCTAGTATATACC 207
```

RESULT 6 425 bp mRNA linear EST 16-OCT-2001
LOCUS BB750925
DEFINITION BB750925 RIKEN full-length enriched, pooled tissues, cerebellum,
etc. Mus musculus cDNA clone G130203B11 3', mRNA sequence.

ACCESSION BB750925.1 GI:16155161
KEYWORDS EST.
SOURCE house mouse.

ORGANISM

Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 425)

REFERENCE
AUTHORS Akimura,T., Arikawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Inocenti,K., Ishii,
Y., Ito,M., Kawai,J., Kojima,Y., Kono,H., Konda,M., Matsuyama,T.,
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,
A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,
Watanishi,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)

JOURNAL Unpublished (2001)
COMMENT Contact: Yoshinori Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,
M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Itawa,M., Ohara,E.,
Watanishi,K., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)

Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,
Y. and Hayashizaki,Y.

Computer-based method for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.

FEATURES

Source

Location/Qualifiers

```
1. 425
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="G130203B11"
/clone_lib="RIKEN full-length enriched, pooled tissues,
cerebellum, etc."
/notes="pooled tissues: (tissue_type=cerebellum,
dev_stage=16 days neonate, sex=mixed),
(tissue_type=cerebellum, dev_stage=0 day neonate,
sex=mixed), (tissue_type=hippocampus, dev_stage=adult,
sex=male), (tissue_type=whole body, dev_stage=9 days
embryo, sex=mixed)" (tissue_type=lung, dev_stage=13 days
embryo, sex=mixed)"
```

BASE COUNT 128 a 109 c 87 g 101 t
ORIGIN
Query Match 24.0%; Score 146; DB 10; Length 425;
Best Local Similarity 100.0%; Pred. No. 1.9e-65;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 411 CAGTTACGACGAGTTTATGCTGCGACCCCTACACACACACTTGTCCAGCCCCAC 470
DB 73 CAGTTACGACGAGTTTATGCTGCGACCCCTACACACACACTTGTCCAGCCCCAC 132
OY 471 CTACGCGCTTGTGTCATGATGCTTTGGCCCTTGAACGATGCGACGATGAGGACA 530
DB 133 CTACGCGCTTGTGTCATGATGCTTTGGCCCTTGAACGATGCGACGATGAGGACA 192
OY 531 TCCTGATGATGTGGTCTGCTCTTT 556
DB 193 TCCTGATGATGTGGTCTGCTCTTT 218
```

RESULT 7

LOCUS

AA818804/c

DEFINITION AA818804 538 bp mRNA linear EST 03-JUL-1999
UT-R-A0-ar-b-11-0-UT.s1 UT-R-A0 Rattus norvegicus cDNA clone
a81c05.r1 Soares NHMpu SI Homo sapiens cDNA clone 1048424 5',
mRNA sequence.

ACCESSION AA818804
VERSION AA818804.1 GI:4227480

KEYWORDS

SOURCE

ORGANISM Rattus norvegicus

source

1. .430
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BG1-aie-e-06-0-UI"
/clone_lib="NIH_BMAP_MSC_N"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The NIH_BMAP_MSC_N library is a normalized library constructed from mouse spinal cord. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Willer Laboratories.
TAG_LIB=NIH_BMAP_MSC_N
TAG_TISSUE=spinal-cord
TAG_SEQ=TCGAA"

BASE COUNT 97 a 85 c 104 g 144 t

ORIGIN

Query Match 27.9%; Score 170; DB 9; Length 430;
Best Local Similarity 100.0%; Pred. No. 4.8e-78;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 411 CAGTTAGGAGAGAGTTTGTGCGGACCCCTACACACACACACTGTGCGAGCCCGAC 470
Db 370 CAGTTAGGAGAGAGTTTGTGCGGACCCCTACACACACACACTGTGCGAGCCCGAC 311

QY 471 CTACGGCGTTGTGCGCATGATGCTTTTGCGCCCTTGACCGATGCCAAGACTAGAGGCA 530
Db 310 CTACGGCGTTGTGCGCATGATGCTTTTGCGCCCTTGACCGATGCCAAGACTAGAGGCA 251

QY 531 TGCATGATATGTGGTCTGCTCTTTCTTTTCATTGACAGGCTAGTATATACC 580
Db 250 TGCATGATATGTGGTCTGCTCTTTCTTTTCATTGACAGGCTAGTATATACC 201

RESULT 4
BF549922/c 533 bp mRNA linear EST 12-DEC-2000
LOCUS BF549922
DEFINITION UI-R-E0-bv-c-04-0-UI.r1 UI-R-E0 Rattus norvegicus cDNA clone
ACCESSION BF549922
VERSION BF549922.1 GI:11659652
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 533)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)

JOURNAL MEDLINE
COMMENT 97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: meoares@blue.weeg.uiowa.edu
cDNA Library Preparation: M.B. Soares Lab Clone distribution:
Clones will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.G.E. Consortium at
LNLN (info@image.llnl.gov). IMAGE ID= 1777166
Seq primer: M13 forward.
Location/Qualifiers
1. .533

FEATURES
source

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-E0-bv-c-04-0-UI"
/clone_lib="UI-R-E0"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: NotI; Site 2: EcoRI; This library consists of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."

BASE COUNT 134 a 125 c 164 g 109 t 1 others

ORIGIN

Query Match 26.3%; Score 160; DB 12; Length 533;
Best Local Similarity 99.5%; Pred. No. 9.1e-73;
Matches 210; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 370 TACGCCAGCCACCCCTGCGCACTGCTGCTCTACAGTGACAGTTTACGAGAGTTAT 429
Db 268 TACGCCAGCCACCCCTGCGCACTGCTGCTCTACAGTGACAGTTTACGAGAGTTAT 209

QY 430 GCTGCCAGCCCTTACACACACACTTGTCTCAGACCCCACTACGGCGTTGTGCAATG 489
Db 208 GCTGCCAGCCCTTACACACACACTTGTCTCAGACCCCACTACGGCGTTGTGCAATG 149

QY 490 AATGCTTTTGCGCCCTTGACCGATGCCAAGACTAGAGCCATGCTATGATGTGGGTCTC 549
Db 148 AATGCTTTTGCGCCCTTGACCGATGCCAAGACTAGAGCCATGCTATGATGTGGGTCTC 89

QY 550 GTTCTTTCTTTCATTGACAGGCTAGTATATACC 580
Db 88 GTTCTTTCTTTCATTGACAGGCTAGTATATACC 58

RESULT 5
A1835011/c 361 bp mRNA linear EST 14-JUN-1999
LOCUS A1835011
DEFINITION UI-M-AM1-a1w-h-01-0-UI.s1 NIH_BMAP_MAM_N Mus musculus cDNA clone
ACCESSION A1835011
VERSION A1835011.1 GI:5469224
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 361)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)

JOURNAL MEDLINE
COMMENT 97044477
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mestr@mail.nih.gov
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized amygdala library cDNA library Preparation: M.B. Soares Lab Clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be

and the oligo-dT track served to verify it as a clone from the normalized hippocampus library cDNA Library Preparation: M.B. Soares Lab Clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined. The following repetitive elements were found in this cDNA sequence: 117-163, >(CAA)n#simple_repeat
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

SOURCE

location/Qualifiers
1..379
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-AQ1-adz-f-10-0-UI"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRTT3D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; The NIH BMAP MHI N library is a normalized library constructed from mouse hippocampus. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories.
TAG_LIB=NIH BMAP_MHI_N
TAG_TISSUE=hippocampus
TAG_SEQ="TTGCA"
BASE COUNT 82 a 75 c 92 g 130 t
ORIGIN

Query Match 29.2%; Score 178; DB 9; Length 379;

Best Local Similarity 100.0%; Pred. No. 2.9e-82;
Matches 178; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

403 TACAGTACAGTTACGAGAGTTATGCTGCGACCCCTACACACACACTGTGCA 462
|||
379 TACAGTACAGTTACGAGAGTTATGCTGCGACCCCTACACACACACTGTGCA 320
|||
463 GCCCCACCTACGCGCTTGTCATGAATGCTTTGCGCCCTTGACCGATGCAAGCT 522
|||
DB 319 GCCCCACCTACGCGCTTGTCATGAATGCTTTGCGCCCTTGACCGATGCAAGCT 260
|||
QY 523 AGAGACGACGTCATGATGTGGTCTGCTTTCTTCTTCAATGAGAGCTGATATACC 580
|||
DB 259 AGAGACGACGTCATGATGTGGTCTGCTTTCTTCTTCAATGAGAGCTGATATACC 202
|||

RESULT 2

A1175239/c

LOCUS A1175239 266 bp mRNA linear EST 20-JAN-1999
DEFINITION EST218774 Normalized rat muscle, Bento Soares Rattus sp. cDNA clone
RMUCE70 3' end, mRNA sequence.

ACCESSION

A1175239

VERSION

A1175239.1

KEYWORDS

EST.

SOURCE

Rattus sp.

ORGANISM

Rattus sp.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 266)
Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,
Kerlavage, A.R. and Adams, M.D.
Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat
Gene Index
Unpublished (1998)
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@igf.org
Seq primer: M13-21.

FEATURES

SOURCE

location/Qualifiers
1..266
/organism="Rattus sp."
/db_xref="ATCC (inhost):2032297"
/db_xref="taxon:10118"
/clone="RMUCE70"
/note="Organ: muscle; Vector: pRTT3D-Pac; Site_1: EcoRI;
Site_2: NotI"
BASE COUNT 65 a 51 c 72 g 78 t
ORIGIN

Query Match 27.9%; Score 170; DB 9; Length 266;

Best Local Similarity 100.0%; Pred. No. 4.1e-78;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 411 CAGTACGACGAGTTATGCTGCGACCCCTACACACACACTGTGCTGAGCCCCGAC 470
|||
DB 246 CAGTACGACGAGTTATGCTGCGACCCCTACACACACACTGTGCTGAGCCCCGAC 187
|||
QY 471 CTACGCGCTTGTCATGAATGCTTTGCGCGCTTGACCGATGCAAGACTGAGAGCA 530
|||
DB 186 CTACGCGCTTGTCATGAATGCTTTGCGCGCTTGACCGATGCAAGACTGAGAGCA 127
|||
QY 531 TGCTATGATGTGGTCTGCTTTCTTCTTCAATGAGAGCTGATATACC 580
|||
DB 126 TGCTATGATGTGGTCTGCTTTCTTCTTCAATGAGAGCTGATATACC 77
|||

RESULT 3

A1850237/c

LOCUS A1850237 430 bp mRNA linear EST 15-JUL-1999
DEFINITION UI-M-BG1-aie-e-06-0-UI.s1 NIH BMAP MSC.N Mus musculus cDNA clone
UI-M-BG1-aie-e-06-0-UI 3', mRNA sequence.

ACCESSION

A1850237

VERSION

A1850237.1

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mestr@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized spinal cord library cDNA Library Preparation: M.B.
Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
available by the means that is soon to be determined. When NIH
determines the means for distribution of the BMAP cDNA clones, this
record will be updated accordingly when that means is determined.
The following repetitive elements were found in this cDNA sequence:
116-162, >(CAA)n#simple_repeat
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

location/Qualifiers

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: March 15, 2003, 15:59:00 ; Search time 856.787 Seconds

(without alignments)
11511.672 Million cell updates/sec

Title: US-09-809-545a-1_COPY_535_1143

Perfect score: 609
Sequence: 1 atgactataataaaagcgcg.....acaaccgttcgtccacat 609

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST.*
1: em_estdb.*
2: em_esthum.*
3: em_estlin.*
4: em_estlm.*
5: em_estloy.*
6: em_estlpl.*
7: em_estlro.*
8: em_estlc.*
9: gb_estl.*
10: gb_estl2.*
11: gb_estl3.*
12: gb_estl4.*
13: gb_estl5.*
14: gb_estl6.*
15: em_estfun.*
16: em_estlom.*
17: gb_estl7.*
18: em_gse_hum.*
19: em_gse_inv.*
20: em_gse_pln.*
21: em_gse_vrt.*
22: em_gse_fun.*
23: em_gse_mam.*
24: em_gse_mus.*
25: em_gse_other.*
26: em_gse_pro.*
27: em_gse_rnd.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	178	29.2	379	9 A1843387 UI-M-AQ1-
2	170	27.9	266	9 A1175239 EST218774
3	170	27.9	430	9 A1850237 A1850237 UI-M-BG1-
4	160	26.3	533	12 BF549922 UI-R-B0-b
5	152	25.0	361	9 A1835011 A1835011 UI-M-AM1-
6	146	24.0	425	10 BB750925 BB750925

Result No.	Score	Query Match Length	ID	Description
7	92	15.1	538	9 AA818804 AA818804 UI-R-A0-a
8	92	15.1	600	13 B1988957 B1988957 4023-45 M
9	80	13.1	408	12 BG795922 BG795922 UTSW_SM38
10	70	11.5	228	14 Z28900 Z28900 HSB87F031 S
11	70	11.5	327	12 BG186689 BG186689 RST5664 A
12	70	11.5	363	14 D60520 D60520 HDM14E06A
13	70	11.5	367	9 A1095813 A1095813 qd20911.x
14	70	11.5	367	9 A1189407 A1189407 qd01h05.x
15	70	11.5	331	9 A1186273 A1186273 qd20b08.x
16	70	11.5	392	9 AA975235 AA975235 cq36c08.s
17	70	11.5	420	9 A1655094 A1655094 wbe7a10.x
18	70	11.5	437	12 BP233478 BP233478 7g33e01.x
19	70	11.5	439	9 A1968176 A1968176 wu14a06.x
20	70	11.5	439	10 AM589795 AM589795 hg22d04.x
21	70	11.5	439	10 BE501380 BE501380 7a41h09.x
22	70	11.5	443	9 AA773715 AA773715 af81c05.x
23	70	11.5	449	9 A1374891 A1374891 ca60912.x
24	70	11.5	466	9 A1917298 A1917298 t896d09.x
25	70	11.5	459	9 A1244212 A1244212 qj86d11.x
26	70	11.5	477	10 AM197589 AM197589 xm44g02.x
27	70	11.5	480	9 A1401040 A1401040 lh27a12.x
28	70	11.5	495	9 A1656926 A1656926 ct48e08.x
29	70	11.5	607	9 AU145000 AU145000 AU145000 AU145000
30	70	11.5	773	10 AV729198 AV729198 AV729198 AV729198
31	70	11.5	774	10 AV729057 AV729057 AV729057 AV729057
32	70	11.5	4039	11 BC026312 BC026312 Homo_sapi
33	70	11.5	283	12 BG185609 BG185609 RST4560 A
34	62	10.2	674	12 BG793800 BG793800 UTSW_SM16
35	54	8.9	218	9 A1189257 A1189257 qc99h03.x
36	54	8.9	261	14 F21037 F21037 HSPD0503 H
37	54	8.9	372	9 A1799929 A1799929 wc41a05.x
38	49	8.0	246	9 A1351484 A1351484 q105e06.x
39	48	7.9	521	13 BM183349 BM183349 fv10e03.x
40	48	7.9	723	12 BG306387 BG306387 fm58d11.x
41	43	7.1	302	14 BM930547 BM930547 UI-E-EJ1-
42	43	7.1	347	9 A1041587 A1041587 ox56g05.x
43	43	7.1	428	9 AA460940 AA460940 zx61d04.s
44	43	7.1	449	9 A1335996 A1335996 qt42e01.x
45	43	7.1	451	9 A1193843 A1193843 qe72h11.x

ALIGNMENTS

RESULT 1
A1843387/c 379 bp mRNA linear EST 14-JUN-1999
LOCUS A1843387
DEFINITION UI-M-AQ1-act-f-10-0-UI-s1 BMAP_MHI N Mus musculus CDNA clone
UI-M-AQ1-act-f-10-0-UI 3', mRNA sequence.
ACCESSION A1843387
VERSION A1843387.1 GI:5477600
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 379)
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: MEST@nhi.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a donald poly A
tail. The sequence tag present in the CDNA between the NotI site

THIS PAGE BLANK (USPTO)

PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 12181; 2537bp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH3166 to AAH1628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13632 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 4684 BP; 1247 A; 1039 C; 1054 G; 1344 T; 0 other;
Query Match 3.3%; Score 20; DB 22; Length 4684;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 154 CTTGTATATACTTCTGCAAT 173
Db 3962 CTTGTATATACTTCTGCAAT 3981
RESULT 15
ABV23885
ID ABV23885 standard; cDNA; 4699 BP.
XX
AC ABV23885;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 23876.
XX
KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KM pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-21314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.

XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 4406-4407; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 4699 BP; 1250 A; 1040 C; 1057 G; 1347 T; 5 other;
Query Match 3.3%; Score 20; DB 23; Length 4699;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 154 CTTGTATATACTTCTGCAAT 173
Db 3977 CTTGTATATACTTCTGCAAT 3996

Search completed: March 15, 2003, 16:12:04
Job time : 157.485 secs

XX (HYSE-) HYSEQ INC.
PA
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Gao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue A, Yang Y, Wehrman T, Goodrich R;
XX
XX WPI: 2001-476283/51.
DR P-PSDB; AAM80096.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX Claim 1; Page 4968; 6221pp; English.
PS
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM79323-AAM80020) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX Sequence 1164 BP; 273 A; 280 C; 325 G; 286 T; 0 other;
SQ
Query Match 7.1%; Score 43; DB 22; Length 1164;
Best Local Similarity 100.0%; Pred. No. 8.9e-12;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 100 CTGTTGTGCCAGGCCAACCCAGGAGATCTTCATGTACAGTG 142
Db 173 CTGTTGTGCCAGGCCAACCCAGGAGATCTTCATGTACAGTG 131
RESULT 13
AAS93634
ID AAS93634 standard; cDNA; 2118 BP.
XX
XX AAS93634;
AC
XX
DT 13-FEB-2002 (first entry)
XX
XX
DE DNA encoding novel human diagnostic protein #29438.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX
XX PR 31-MAR-2000; 2000US-0540217.
XX
XX PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
PI
XX
DR WPI: 2001-639362/73.
DR P-PSDB; ABG29447.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID NO 29438; 103pp; English.
PS
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAK54197-AAK94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2118 BP; 557 A; 516 C; 542 G; 503 T; 0 other;
SQ
Query Match 6.4%; Score 39; DB 23; Length 2118;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 412 AGTTACGACGAGTTATGTCGCCAGCCCTACACACAC 450
Db 1300 AGTTACGACGAGTTATGTCGCCAGCCCTACACACAC 1338
RESULT 14
AAH14583
ID AAH14583 standard; cDNA; 4684 BP.
XX
XX AAH14583;
AC
XX
XX 26-JUN-2001 (first entry)
DT
XX
XX
DE Human cDNA sequence SEQ ID NO:12181.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
OS
XX EP1074617-A2.
XX
XX PD 07-FEB-2001.
XX
XX PF 28-JUL-2000; 2000EP-0116126.
XX
XX PR 29-JUL-1999; 99JP-0248036.
XX
XX PR 27-AUG-1999; 99JP-0300253.
XX
XX PR 11-JAN-2000; 2000JP-0118776.
XX
XX PR 02-MAY-2000; 2000JP-0183767.
XX
XX PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
PA
XX Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602

XX The present invention describes oligonucleotide libraries for detecting
CC messenger RNAs that populate a (sub-)transcriptome, where the
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises
CC several oligonucleotides, each capable of hybridizing selectively to a
CC set of messenger RNAs transcribed from a given transcription unit of
CC the genome, which encodes one or more messenger RNA splice variants.
CC The oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterizing the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
CC transcripts. The libraries may also be used as specialised mini
CC libraries to detect transcripts of a sub-transcriptome under a
CC particular biological or pathological state, and so allowing the
CC detection of tissue- and pathology-specific genes such as those genes
CC only expressed in specific tissue under a specific pathological
CC condition, to detect developmental specific genes, and to detect RNA
CC transcripts and splice variants of a transcriptome of a patient suffering
CC from a particular disorder. ABN27253 to ABN59589 represent
CC oligonucleotide sequences from rats, humans and mice, which are used in
CC the exemplification of the present invention.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WFO
CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).
XX

Sequence 60 BP; 11 A; 14 C; 16 G; 19 T; 0 other;

Query Match 7.9%; Score 48; DB 24; Length 60;
Best Local Similarity 100.0%; Pred. No. 2.4e-14; Indels 0; Gaps 0;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 511 GATGCCAAGACTAGAGCCATGCTGATGATGAGTGGTCTGTTCTTCT 558
DB 13 GATGCCAAGACTAGAGCCATGCTGATGATGAGTGGTCTGTTCTTCT 660

RESULT 11
ABK34602
ID ABK34602 standard; cDNA; 3189 BP.

XX AC ABK34602;

XX DT 08-MAY-2002 (first entry)

XX DE Human cDNA for novel secreted protein, SEQ ID 371.

XX Human; 69; gene; secreted protein; immune deficiency; viral infection;
KW bacterial infection; fungal infection; autoimmune disorder; burn;
KW rheumatoid arthritis; multiple sclerosis; autoimmune chryoiditis;
KW diabetes; allergy; asthma; neurodegenerative disease; Crohn's disease;
KW Alzheimer's disease; Parkinson's disease; liver fibrosis; tumour;
KW coagulation disorder; haemophilia; inflammatory disorder; ulcer;
KW tissue regeneration; wound healing; haematopoiesis; myeloid deficiency;
KW lymphoid cell deficiency.

XX KM Homo sapiens.

XX OS Homo sapiens.

XX PN MO2001.77290-A2.

XX PD 18-OCT-2001.

XX PF 29-MAR-2001; 2001WO-US10295.

XX PR 06-APR-2000; 2000US-194941P.

XX PA (GEMV) GENETICS INST INC.

XX PI Wong GG, Clark HF, Fachtel K, Agostino MJ, Howes SH, Resnick RJ;

XX DR Gulukota K, Graham JR;

XX WPI; 2002-179323/23.

PT Six hundred and twenty five polynucleotides derived from a variety of
PT human tissue sources which encode secreted proteins, useful for
PT treating immune deficiencies and disorders such as autoimmune disorders
PT -
XX
PS Claim 1; Page 188-189; 339pp; English.

XX The invention relates to 625 polynucleotides which have been derived from
CC a variety of human tissue sources and which encode novel secreted
CC proteins, their complements and sequences that hybridise to them.
CC Also included are a vector comprising the polynucleotide, a host cell
CC transformed with the vector, the proteins encoded by the
CC polynucleotides, antibodies that bind to the proteins and identification
CC of modulators of the proteins or the expression of the polynucleotide.
CC The polynucleotides can be used as probes for the identification
CC and isolation of full length cDNA and genomic DNA. The polynucleotides
CC and proteins can also be used as nutritional supplements. The protein
CC is useful in the treatment of various immune deficiencies and disorders
CC such as viral infections, bacterial infections, fungal infections,
CC autoimmune disorders (e.g. rheumatoid arthritis, multiple sclerosis,
CC autoimmune thyroiditis and diabetes) and allergic reactions and
CC conditions (e.g. asthma). They are also useful for treating
CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
CC useful for tissue regeneration, for wound healing and in the treatment
CC of burns, incisions and ulcers. The proteins are also useful for
CC regulating haematopoiesis, for treating myeloid or lymphoid cell
CC deficiencies. The present sequence is one of the 625 cDNA sequences
CC encoding a secreted protein.

Sequence 3189 BP; 1005 A; 554 C; 631 G; 999 T; 0 other;

Query Match 7.7%; Score 47; DB 24; Length 3189;
Best Local Similarity 100.0%; Pred. No. 8.1e-14; Indels 0; Gaps 0;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 534 TGATGATGGGCTCGTCTTTCTTCTTCTGACGCTAGATATATACC 580
DB 515 TGATGATGGGCTCGTCTTTCTTCTTCTGACGCTAGATATATACC 561

RESULT 12
AAK53229/C
ID AAK53229 standard; cDNA; 1164 BP.

XX AC AAK53229;

XX DT 06-NOV-2001 (first entry)

XX DE Human polynucleotide SEQ ID NO 2758.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.

XX KM Homo sapiens.

XX OS Homo sapiens.

XX PN MO2001.57190-A2.

XX PD 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-US04098.

XX PR 03-FEB-2000; 2000US-0496914.

XX PR 27-APR-2000; 2000US-0560875.

XX PR 20-JUN-2000; 2000US-0598075.

XX PR 19-JUL-2000; 2000US-0620325.

XX PR 01-SEP-2000; 2000US-0654936.

XX PR 15-SEP-2000; 2000US-0663561.

XX PR 20-OCT-2000; 2000US-0693325.

XX PR 30-NOV-2000; 2000US-0728422.

Query Match 11.5%; Score 70; DB 22; Length 607;
Best Local Similarity 100.0%; Pred. No. 1.3e-25;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 511 GATGCCAAGACTAGAGCCATGATGATGGGCTCGTTCTTTCTTTCATTGCAGGCT 570
DB 201 GATGCCAAGACTAGAGCCATGATGATGGGCTCGTTCTTTCTTTCATTGCAGGCT 142

QY 571 AGTATATACC 580
DB 141 AGTATATACC 132

RESULT 7
AAH13824
ID AAH13824 standard; cDNA; 1513 BP.
XX
AC AAH13824;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:10786.
XX
KM Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN BP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Oca T, Isogai T, Nishikawa T, Hayashi K, Saio K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
PS Claim 8; SEQ ID 10786; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNA easily without any specialised methods. AAH03466 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AA992446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 1513 BP; 354 A; 418 C; 395 G; 346 T; 0 other;

Query Match 11.5%; Score 70; DB 22; Length 1513;
Best Local Similarity 100.0%; Pred. No. 1.4e-25;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 511 GATGCCAAGACTAGAGCCATGATGATGGGCTCGTTCTTTCTTTCATTGCAGGCT 570
DB 1313 GATGCCAAGACTAGAGCCATGATGATGGGCTCGTTCTTTCTTTCATTGCAGGCT 1312

QY 571 AGTATATACC 580
DB 1373 AGTATATACC 1382

RESULT 8
AAK52245
ID AAK52245 standard; cDNA; 1800 BP.
XX
AC AAK52245;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO. 790.
XX
KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KM tissue growth factor; immunomodulatory; cancer; leukaemia;
KM nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Gao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AD, Yang Y, Wehrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
XX
PT P-PSDB; AAM79112.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX Claim 1; Page 2643-2645; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK51435) and the
XX encoded polypeptides (AAM78333-AAK80302) that exhibit activity relating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polypeptides and polynucleotides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or

ABK34530/C
ID ABK34530 standard; cDNA; 539 BP.
XX
AC ABK34530;
XX
DT 08-MAY-2002 (first entry)
XX
DE Human cDNA for novel secreted protein, SEQ ID 299.
XX
KM Human; ss; gene; secreted protein; immune deficiency; viral infection;
KM bacterial infection; fungal infection; autoimmune disorder; burn;
KM rheumatoid arthritis; multiple sclerosis; autoimmune thyroiditis;
KM diabetes; allergy; asthma; neurodegenerative disease; Crohn's disease;
KM Alzheimer's disease; Parkinson's disease; liver fibrosis; tumour;
KM coagulation disorder; haemophilia; inflammatory disorder; ulcer;
KM tissue regeneration; wound healing; haematopoiesis; myeloid deficiency;
KM lymphoid cell deficiency.
XX
OS Homo sapiens.
XX
PN MO200177290-A2.
XX
PD 18-OCT-2001.
XX
PF 29-MAR-2001; 2001MO-US10295.
XX
PR 06-APR-2000; 2000US-194941P.
XX
PA (GENY) GENETICS INST INC.
XX
PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
PI Gulkoka K, Graham JR;
XX
DR WPI: 2002-179323/23.
XX
PT Six hundred and twenty five polynucleotides derived from a variety of
PT human tissue sources which encode secreted proteins, useful for
PT treating immune deficiencies and disorders such as autoimmune disorders
PT
XX
XX Claim 1; Page 153; 339pp; English.
XX
PS The invention relates to 625 polynucleotides which have been derived from
XX a variety of human tissue sources and which encode novel secreted
XX proteins, their complements and sequences that hybridise to them.
XX Also included are a vector comprising the polynucleotide, a host cell
XX transformed with the vector, the proteins encoded by the
XX polynucleotides, antibodies that bind to the proteins and identification
XX of modulators of the proteins or the expression of the polynucleotide.
XX The polynucleotides can be used as probes for the identification
XX and isolation of full length cDNA and genomic DNA. The polynucleotides
XX and proteins can also be used as nutritional supplements. The protein
XX is useful in the treatment of various immune deficiencies and disorders
XX such as viral infections, bacterial infections, fungal infections,
XX autoimmune disorders (e.g. rheumatoid arthritis, multiple sclerosis,
XX autoimmune thyroiditis and diabetes) and allergic reactions and
XX conditions (e.g. asthma). They are also useful for treating
XX neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
XX disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
XX inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
XX useful for tissue regeneration, for wound healing and in the treatment
XX of burns, incisions and ulcers. The proteins are also useful for
XX regulating haematopoiesis, for treating myeloid or lymphoid cell
XX deficiencies. The present sequence is one of the 625 cDNA sequences
XX encoding a secreted protein.
XX
SQ Sequence 539 BP; 132 A; 132 C; 156 G; 119 T; 0 other;

Query Match 11.5%; Score 70; DB 24; Length 539;
Best Local Similarity 100.0%; Pred. No. 1.3e-25;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

511 GATGCCAAGACTAGAGCCATGCTGATGTGGTCTCGTCTTCTTCATTCAGAGCT 570

Db 191 GATGCCAAGACTAGAGCCATGCTGATGTGGTCTCGTCTTCTTCATTCAGAGCT 132
Oy 571 AGTATATACC 580
Db 131 AGTATATACC 122

RESULT 6
AAH09205/C
ID AAH09205 standard; cDNA; 607 BP.
XX
AC AAH09205;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (3'-primer) SEQ ID NO:6040.
XX
KM Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI: 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 3; SEQ ID 6040; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesising 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesising polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
SQ Sequence 607 BP; 140 A; 152 C; 179 G; 129 T; 7 other;

XX 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #29436.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
OS WO200175067-A2.
PN 11-OCT-2001.
PD 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
XX P-PSDB; ABG29445.
DR New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 1; SEQ ID No 29436; 103bp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridization probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 406 BP; 94 A; 97 C; 85 G; 102 T; 28 other;

Query Match 11.5%; Score 70; DB 23; Length 406;
Best Local Similarity 100.0%; Pred. No. 1.3e-25;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 511 GATGCAAGACTAGAGCCATGCTGATGATGCGTCTGTTCTTTCATTGACAGCT 570
DB 142 GATGCAAGACTAGAGCCATGCTGATGATGCGTCTGTTCTTTCATTGACAGCT 201
QY 571 AGTATATACC 580
DB 202 AGTATATACC 211

RESULT 4
AAS93633
ID AAS93633 standard; cDNA; 481 BP.

XX AAS93633;
AC 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #29437.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
OS WO200175067-A2.
PN 11-OCT-2001.
PD 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
XX P-PSDB; ABG29446.
DR New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 1; SEQ ID No 29437; 103bp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 481 BP; 100 A; 145 C; 123 G; 113 T; 0 other;

Query Match 11.5%; Score 70; DB 23; Length 481;
Best Local Similarity 100.0%; Pred. No. 1.3e-25;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 511 GATGCAAGACTAGAGCCATGCTGATGATGCGTCTGTTCTTTCATTGACAGCT 570
DB 306 GATGCAAGACTAGAGCCATGCTGATGATGCGTCTGTTCTTTCATTGACAGCT 365
QY 571 AGTATATACC 580
DB 366 AGTATATACC 375

RESULT 5

PA (SCIO-) SCIOS INC.
 XX Stanton LW, White RT;
 XX WPI: 2002-010779/01.
 DR P-PSDB; AAU70146.
 XX Novel secreted factor polypeptide useful for treating cardiac diseases
 PT such as arteriosclerosis, myocardial infarction, inflammatory diseases
 XX such as asthma, stroke, and rheumatoid arthritis and renal diseases
 PS Claim 1: Fig 1: 189pp; English.
 XX The invention relates to rat secreted factor polypeptides and the
 CC polynucleotides encoding them. The sequences are useful for treating
 CC cardiac, renal or inflammatory diseases. These include cardiac diseases
 CC such as congestive heart failure, myocarditis, dilated congestive
 CC cardiomyopathy, angina pectoris, myocardial infarction, cardiac
 CC arrhythmia, pulmonary hypertension, arteriosclerosis, atherosclerosis and
 CC cardiac tumours, renal diseases such as glomerulonephritis, nephrotic
 CC syndrome, renal infarction, hereditary nephritis, polycystic kidney
 CC disease, chronic renal failure, renal vein thrombosis and medullary
 CC sponge kidney and inflammatory diseases such as asthma, rheumatoid
 CC arthritis, osteoarthritis, stroke, psoriasis, restenosis, graft versus
 CC host reaction, Crohn's disease, ulcerative colitis and Alzheimer's
 CC disease. Sequences AAS94693-AAS94745 represent cDNA clones, which encode
 CC the secreted factor polypeptides of the invention, and oligonucleotide
 CC probes and PCR primers.
 XX
 XX Sequence 1340 BP, 366 A, 394 C, 317 G, 263 T, 0 other;
 SQ
 Query Match 100.0%; Score 609; DB 24; Length 1340;
 Best Local Similarity 100.0%; Pred. No. 1.7e-30;
 Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACTAATAAAGGCGCTGAACCCCTACCAATGCTGGAATTAATCAAGTTGG 60
 DB 535 ATGACTAATAAAGGCGCTGAACCCCTACCAATGCTGGAATTAATCAAGTTGG 594
 QY 61 GGGGGGCTTACAGCCCGGACCTTATGACAGGCGAGGTGTGTGCGACAGGCAACAG 120
 DB 595 GGGGGGCTTACAGCCCGGACCTTATGACAGGCGAGGTGTGTGCGACAGGCAACAG 654
 QY 121 GAGGATCTTCCATGATACAGTGGCCCGAGTTCACTTGTATATATCTTCAATGCTTGGC 180
 DB 655 GAGGATCTTCCATGATACAGTGGCCCGAGTTCACTTGTATATATCTTCAATGCTTGGC 714
 QY 181 TTTCCATATCCGGCCCGCACTGTGCAAGTGCATACCGAGGGGCTACCTTGAAGGCGT 240
 DB 715 TTTCCATATCCGGCCCGCACTGTGCAAGTGCATACCGAGGGGCTACCTTGAAGGCGT 774
 QY 241 GGTCCGACCGGTGACAAACCTTCAGAGCTGGGGGCGCCCGCCCAATCCCGGCTAT 300
 DB 775 GGTCCGACCGGTGACAAACCTTCAGAGCTGGGGGCGCCCGCCCAATCCCGGCTAT 834
 QY 301 GCGGAGTGTGTATCAAGAGCCAGTGTATGCAATTAATTTGTACAGGGTGTACGCT 360
 DB 835 GCGGAGTGTGTATCAAGAGCCAGTGTATGCAATTAATTTGTACAGGGTGTACGCT 894
 QY 361 GCATACCGCTAGCCGACCCGACCTGCTGCTGCTGCTTACAGTACAGTACGGA 420
 DB 895 GCATACCGCTAGCCGACCCGACCTGCTGCTGCTGCTTACAGTACAGTACGGA 954
 QY 421 CGAGTTATGCTGCGGACCCCTACCAACACTTGTCTCAGCCCGCCCACTTACGGGCTT 480
 DB 955 CGAGTTATGCTGCGGACCCCTACCAACACTTGTCTCAGCCCGCCCACTTACGGGCTT 1014
 QY 481 GGTGCATGAATGCTTTTGGCCCTTACCGATGCCAAGAATAAGATGCTGATGAT 540
 DB 1015 GGTGCATGAATGCTTTTGGCCCTTACCGATGCCAAGAATAAGATGCTGATGAT 1074
 QY 541 GTGGGTCTGTTTCTTCTTATGCAAGGCTAGTATATACCAAGGGGATTCACCGTTT 600

DB 1075 GTGGGTCTGTTTCTTCTTATGCAAGGCTAGTATATACCAAGGGGATACACGTTT 1134
 QY 601 GGTCCATAT 609
 DB 1135 GGTCCATAT 1143
 RESULT 2
 ID AAK54001/c
 AA AAK54001 standard; cDNA; 327 BP.
 XX
 XX AAK54001;
 XX
 XX 16-NOV-2001 (first entry)
 DT
 DE Murine transcription associated protein encoding cDNA SEQ ID 566.
 XX
 XX Murine; liver; gene library; amino acid synthesis; binding protein;
 KW cell metabolism; energy metabolism; fatty acid metabolism; synthesis;
 KW phospholipid metabolism; purine; pyrimidine; nucleoside; nucleotide;
 KW replication; transcription; translation; transport protein; ss.
 XX
 OS Mus musculus.
 XX
 XX DE20103510-UI.
 XX
 XX 07-JUN-2001.
 XX
 XX 28-FEB-2001; 2001DE-2003510.
 XX
 XX 02-DEC-1999; 99DE-1058160.
 XX
 XX (LION-) LION BIOSCIENCE AG.
 PA
 DR WPI: 2001-368570/39.
 XX
 PT Gene library containing sequences with specific 3'-ends and no polyA
 PT tail, encoding proteins involved in a wide range of cellular processes
 PT
 PS Claim 15; Page 174; 251pp; German.
 XX
 CC This invention describes a novel gene library (A) comprises a gene
 CC sequence (or its part) encoding a protein involved in amino acid
 CC synthesis, cellular/energy metabolism, metabolism of
 CC fatty acids/phospholipids, synthesis or breakdown of
 CC purines/pyrimidines/nucleosides/nucleotides, DNA
 CC replication/transcription/translation, or is a transport/binding protein.
 CC (A) are produced that correspond to the 3'-end of mRNA but without the
 CC polyA tail. They can be prepared more efficiently and with less effort
 CC than conventional libraries. AAK53436-AAK54275 represent fragments of the
 CC gene library described in the method of the invention.
 XX
 XX Sequence 327 BP, 71 A, 65 C, 84 G, 107 T, 0 other;
 SQ
 Query Match 14.1%; Score 86; DB 22; Length 327;
 Best Local Similarity 100.0%; Pred. No. 8.6e-34;
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 495 TTTTGGCCCTTACCGATGCCAAGACTAGGACCATGTGATGATGGGCTCTGTTCT 554
 DB 272 TTTTGGCCCTTACCGATGCCAAGACTAGGACCATGTGATGATGGGCTCTGTTCT 213
 QY 555 TTCTTCATTCGAGGCTAGATATACC 580
 DB 212 TTCTTCATTCGAGGCTAGATATACC 187

RESULT 3
 ID AAS93632 standard; cDNA; 406 BP.
 XX
 XX AAS93632;
 AC

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 15, 2003, 12:52:03 ; Search time 142.485 Seconds
(without alignments)
9625.317 Million cell updates/sec

Title: US-09-809-545a-1_COPY_535_1143

Perfect score: 1 atgactaataaaagccgcgt.....acaacgcttcctccatcat 609

Sequence: OLIGO NUC

Scoring table: Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 112599159 residues

Word size: 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: N_Geneseq_101002.*

1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	609	100.0	1340	24	AAS94693
2	86	14.1	327	22	AAK54001
3	70	11.5	406	23	AAS93632
4	70	11.5	481	23	AAS93633
5	539	11.5	539	24	ABK34530
6	70	11.5	607	22	AAH09205
7	70	11.5	1513	22	AAH13824
8	70	11.5	1800	22	AAK52245
9	70	11.5	2372	21	AAA07075

10	48	7.9	60	24	ABN34672	Human spliced tran
11	47	7.7	3189	24	ABK34602	Human cDNA for nov
12	43	7.1	1164	22	AAK53229	Human polynucleoti
13	39	6.4	2118	23	AAS93634	DNA encoding novel
14	20	3.3	4684	22	AAH14583	Human cDNA sequenc
15	20	3.3	4699	23	ABV23385	Human prostate exp
16	20	3.3	4699	23	ABV23397	Human prostate exp
17	20	3.3	4699	23	ABV29766	Human prostate exp
18	20	3.3	4699	23	ABV29862	Human prostate exp
19	19	3.1	50	22	AAJ11184	Human SNP oligonuc
20	19	3.1	2547	23	ABL17965	Drosophila melanog
21	19	3.1	2547	23	ABL17965	Drosophila melanog
22	19	3.1	6846	23	ABL18560	Drosophila melanog
23	19	3.1	12631	23	ABL17964	Drosophila melanog
24	18	3.0	332	24	ABL69669	Prostate cancer re
25	18	3.0	372	23	AA571730	DNA encoding novel
26	18	3.0	430	23	AA566448	DNA encoding novel
27	18	3.0	755	23	AA585509	DNA encoding novel
28	18	3.0	1668	23	AA589718	DNA encoding novel
29	18	3.0	1668	23	AA589818	DNA encoding novel
30	18	3.0	2262	23	ABL08285	Drosophila melanog
31	18	3.0	2830	24	AA516455	Corn cDNA encoding
32	18	3.0	3035	23	AA571731	DNA encoding novel
33	18	3.0	3431	22	AA585513	First splice varia
34	18	3.0	3446	24	ABK66136	cDNA encoding huma
35	18	3.0	3468	22	AA585514	Second splice vari
36	18	3.0	3582	22	AA546157	Human DNA encoding
37	18	3.0	3741	23	ABL05203	Drosophila melanog
38	18	3.0	4130	23	ABL08284	Drosophila melanog
39	18	3.0	4830	20	ABA96894	Glucuronidase oxd
40	18	3.0	4830	21	AA583156	DNA encoding Gluco
41	18	3.0	5920	23	ABL05202	Drosophila melanog
42	18	3.0	8159	22	AA199600	Human expressed po
43	18	3.0	8159	22	ABA06824	Human genomic DNA
44	18	3.0	8159	22	AA529006	Genomic sequence #
45	18	3.0	8159	22	AA529758	Human endocrine po

ALIGNMENTS

RESULT 1	
AAS94693	standard; cDNA; 1340 BP.
XX	
XX	AAS94693;
XX	
DT	12-MAR-2002 (first entry)
XX	
DE	Rat secreted factor DNA clone P0184_D11 #1.
XX	
KW	Rat: secreted factor polypeptide; cardiac disease; kidney;
KW	inflammatory disease; congestive heart failure; myocarditis; asthma; ss;
KW	dilated congestive cardiomyopathy; angina pectoris; cardiac arrhythmia;
KW	myocardial infarction; pulmonary hypertension; arteriosclerosis; stroke;
KW	atherosclerosis; cardiac tumour; glomerulonephritis; nephrotic syndrome;
KW	renal infarction; hereditary nephritis; polycystic kidney disease;
KW	chronic renal failure; renal vein thrombosis; medullary sponge kidney;
KW	rheumatoid arthritis; osteoarthritis; psoriasis; reestenosis; PCR primer;
KW	graft versus host reaction; Crohn's disease; ulcerative colitis; probe;
KW	Alzheimer's disease; gene therapy.
XX	
OS	Rattus norvegicus.
XX	
PN	WO200174901-A2.
XX	
PD	11-OCT-2001.
XX	
PF	23-MAR-2001; 2001WO-US09555.
XX	
PR	31-MAR-2000; 2000US-193548P.
XX	
PR	14-MAR-2001; 2001US-0809545.
XX	

REFERENCE
AUTHORS
TITLE
JOURNAL

of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 115860)

Worley, K.C.

Direct Submission

Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On Jul 18, 2002 this sequence version replaced gi:20514522.

COMMENT

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: CXLIC

Center clone name: CH230-34C2

----- Summary Statistics

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap: version 0.990329

Consensus quality: 99693 bases at least Q40

Consensus quality: 101936 bases at least Q20

Consensus quality: 102743 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_drafc_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 21 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1 1347: contig of 1347 bp in length
1348 1447: gap of unknown length
1448 2834: contig of 1387 bp in length
2835 2934: gap of unknown length
2935 4253: contig of 1319 bp in length
4254 4353: gap of unknown length
4354 5477: contig of 1124 bp in length
5478 5577: gap of unknown length
5578 7484: contig of 1907 bp in length
7485 7585: gap of unknown length
7586 9406: contig of 1822 bp in length
9407 9506: gap of unknown length
9507 11712: contig of 2206 bp in length
11713 13462: contig of 1650 bp in length
13463 13562: gap of unknown length
13563 17222: contig of 3660 bp in length
17223 17322: gap of unknown length
17323 20617: contig of 3295 bp in length
20618 20717: gap of unknown length
20718 24825: contig of 4108 bp in length
24826 24925: gap of unknown length
24926 28811: contig of 3886 bp in length
28812 28911: gap of unknown length
28912 33686: contig of 4775 bp in length
33687 33786: gap of unknown length
33787 40217: contig of 6431 bp in length
40218 40317: gap of unknown length
40318 47912: contig of 7595 bp in length
47913 48012: gap of unknown length
48013 56482: contig of 8470 bp in length
56483 56582: gap of unknown length
56583 66204: contig of 9622 bp in length
66205 66304: gap of unknown length
66305 74546: contig of 8242 bp in length
74547 74646: gap of unknown length
74647 86522: contig of 11876 bp in length
86523 86622: gap of unknown length

* 86623 100049: contig of 13427 bp in length
* 100050 100149: gap of unknown length
* 100150 115860: contig of 15711 bp in length.
Location/Qualifiers
1..115860
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-34C2"

FEATURES

SOURCE

BASE COUNT 33039 a 22417 c 22786 g 30574 t 7044 others
ORIGIN

Query Match 11.0%; Score 67; DB 2; Length 115860;
Best Local Similarity 100.0%; Pred. No. 7.9e-28;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 347 AGGGGTGTTACGTCATACCGCTACGCGCCAGCCACCCCTGCACCTGCTGTACA 406
DB 34697 AGGGGTGTTACGTCATACCGCTACGCGCCAGCCACCCCTGCACCTGCTGTACA 34756
QY 407 GTGACAG 413
DB 34757 GTGACAG 34763

Search completed: March 15, 2003, 22:52:37
Job time : 1895.46 secs

repeat_region	complement(70389, .70704)	
repeat_region	/rpt_family="MER1"	
repeat_region	complement(72166, .72421)	
repeat_region	/rpt_family="MER3"	
repeat_region	72501, .72769	
repeat_region	/rpt_family="Alu"	
repeat_region	72843, .72956	
repeat_region	/rpt_family="MERS"	
repeat_region	complement(76303, .76564)	
repeat_region	/rpt_family="Alu"	
repeat_region	97534, .97700	
repeat_region	/rpt_family="MER3"	
repeat_region	complement(98469, .98526)	
repeat_region	/rpt_family="Alu"	
repeat_region	complement(99089, .99381)	
repeat_region	/rpt_family="Alu"	
repeat_region	101648, .101930	
repeat_region	/rpt_family="Alu"	
repeat_region	102750, .103056	
repeat_region	/rpt_family="Alu"	
repeat_region	104010, .104277	
repeat_region	/rpt_family="Alu"	
repeat_region	complement(111121, .111263)	
repeat_region	/rpt_family="MERS"	
repeat_region	complement(112263, .112476)	
repeat_region	/rpt_family="MER30"	
repeat_region	112556, .113234	
repeat_region	/rpt_family="Alu"	
repeat_region	114385, .114657	
repeat_region	/rpt_family="Alu"	
repeat_region	complement(117031, .117194)	
repeat_region	/rpt_family="MER20"	
repeat_region	complement(118464, .118770)	
repeat_region	/rpt_family="Alu"	
repeat_region	119271, .119549	
repeat_region	/rpt_family="Alu"	
repeat_region	complement(122831, .122940)	
repeat_region	/rpt_family="MER41"	
repeat_region	123471, .123742	
repeat_region	/rpt_family="Alu"	
repeat_region	complement(129660, .129972)	
repeat_region	/rpt_family="Alu"	
repeat_region	135335, .135310	
repeat_region	/rpt_family="MLT1"	
repeat_region	135874, .136157	
repeat_region	/rpt_family="Alu"	
repeat_region	139775, .140048	
repeat_region	/rpt_family="Alu"	
repeat_region	141734, .141814	
repeat_region	/rpt_family="MIR"	
repeat_region	143071, .143352	
repeat_region	/rpt_family="Alu"	
repeat_region	148286, .148773	
repeat_region	/rpt_family="MER1"	
repeat_region	150431, .150707	
repeat_region	/rpt_family="Alu"	
repeat_region	155151, .155387	
repeat_region	/rpt_family="Alu"	
repeat_region	complement(156101, .156213)	
repeat_region	/rpt_family="MER20"	
repeat_region	complement(156636, .156920)	
repeat_region	/rpt_family="Alu"	
repeat_region	complement(159633, .159931)	
repeat_region	/rpt_family="Alu"	
repeat_region	complement(160378, .160658)	
repeat_region	/rpt_family="Alu"	
repeat_region	161617, .162140	
repeat_region	/rpt_family="Alu"	
repeat_region	complement(163200, .163315)	
repeat_region	/rpt_family="MIR"	
CASE COUNT	48114 a 34256 c 33738 g 48430 t	
ORIGIN		

Query Match	11.5%	Score 70	DB 9	Length 164538
Best Local Similarity	100.0%	Pred. No. 1.3e-29		
Matches 70	Conservative 0	Mismatches 0	Indels 0	Gaps 0

Qy	511	GATGCACAGACTAGAGCAGCATGCTATGATGTGGGTCTCGTTCTTTCTTCATTTGAGAGCT	570
Db	61328	GATGCCAGACTAGAGCAGCATGCTATGATGTGGGTCTCGTTCTTTCTTCATTTGAGAGCT	61269

Qy	571	AGTATATACC	580
Db	61268	AGTATATACC	61259

RESULT 15					
LOCUS	AC120661				
DEFINITION	AC120661	115860 bp	DNA	linear	HTG-23-JUL-2002
ACCESSION	AC120661.2	GI:21902877			
VERSION	AC120661.2	GI:21902877			
KEYWORDS	HTG; HTGS; PHASE1.				
SOURCE	Rattus norvegicus.				
ORGANISM	Rattus norvegicus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				

REFERENCE	1 (bases 1 to 115860)					
AUTHORS	Munry,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Albrooks,S.L., Amarungu,H.C., Are,J.R., Ayale,M., Banks,T., Barbatoria,J., Benton,J., Bimase,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowe,S., Britva,M., Brown,M., Brown,M., Bryant,N.P., Buluy,C., Burck,P., Burkelt,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dabonne,S.R., David,R., Davila,M.L., Davis,C., Davy-carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Eamhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homel,F., Howard,S., Huber,J., Huijjs,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.J., Li,Z., Lichteage,O., Lieu,C., Liu,J., Liu,W., Loulsged,H., Lozada,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhinney,E., Mcloed,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S., Oguni,M., Okwundu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojoubkan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sison,I., Sodegren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tameris,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Umanu,K., Vasquez,L., Vera,Y., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.					
TITLE	Unpublished					
JOURNAL	2 (bases 1 to 115860)					
REFERENCE	Worley,K.C.					
AUTHORS	Submitted					
JOURNAL	Submitted (09-MAY-2002)					

Human Genome Sequencing Center, Department

BASE COUNT 950 a 775 c 711 g 912 t
ORIGIN
Query Match 11.5%: Score 70; DB 9; Length 3343;
Best Local Similarity 100.0%; Pred. No. 1.4e-29;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 511 GATCCAGACTAGAGCCATGCTGATGAGTGGCTCGTTCTTTCTTCATTCGAGGCT 570
Db 1462 GATCCAGACTAGAGCCATGCTGATGAGTGGCTCGTTCTTTCTTCATTCGAGGCT 1521
Qy 571 AGTATATACC 580
Db 1522 AGTATATACC 1531
RESULT 14 AC005774 164538 bp DNA linear PRI 05-OCT-1998
AC005774/c Locus Homo sapiens chromosome 16, BAC clone 2603 (LANL), complete
DEFINITION sequence.
AC005774 AC005774.1 GI:3702264
VERSION AC005774.1
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 164538)
Ricke,D.O.
TITLE Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
UNPUBLISHED
2 (bases 1 to 164538)
Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,
White,S., Yeng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.
and Deaven,L.
TITLE Sequencing of Human Chromosome 16p13.3
UNPUBLISHED
3 (bases 1 to 164538)
Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,
White,S., Yeng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.
and Deaven,L.
TITLE Direct Submission
Submitted (05-OCT-1998) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
FEATURES
source
1..164538
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p13.3"
/clone="2603"
repeat_region
1378..1665
/rpt_family="Alu"
repeat_region
complement(2144..2320)
/rpt_family="MER21"
repeat_region
complement(2198..2416)
/rpt_family="MER39"
repeat_region
complement(2450..2725)
/rpt_family="Alu"
repeat_region
complement(3157..3393)
/rpt_family="Alu"
repeat_region
complement(3524..3792)
/rpt_family="Alu"
5731..6020
/rpt_family="Alu"
repeat_region
complement(7557..7727)

/rpt_family="MER4"
8282..8928
/rpt_family="LTR8"
repeat_region
10055..10310
/rpt_family="THE1"
repeat_region
10472..10618
/rpt_family="L1"
repeat_region
10639..10939
/rpt_family="Alu"
repeat_region
complement(11109..11357)
/rpt_family="Alu"
12050..12312
/rpt_family="Alu"
repeat_region
complement(14642..14755)
/rpt_family="L1"
17429..18370
/rpt_family="MSTAR"
repeat_region
complement(19470..19819)
/rpt_family="MLT1"
repeat_region
complement(20352..20627)
/rpt_family="Alu"
23808..24058
/rpt_family="Alu"
repeat_region
27654..28217
/rpt_family="L1"
repeat_region
complement(32764..32956)
/rpt_family="L1"
repeat_region
complement(33392..33503)
/rpt_family="Alu"
34586..34870
/rpt_family="Alu"
repeat_region
34936..35489
/rpt_family="MLT2B2"
repeat_region
complement(36940..37172)
/rpt_family="Alu"
37173..37437
/rpt_family="THE1"
repeat_region
41428..41745
/rpt_family="THE1"
repeat_region
complement(42057..43927)
/rpt_family="L1"
43924..44555
/rpt_family="L1"
repeat_region
45128..45367
/rpt_family="Alu"
repeat_region
45467..45704
/rpt_family="Alu"
repeat_region
46192..46578
/rpt_family="MER25"
repeat_region
complement(46779..47141)
/rpt_family="THE1"
repeat_region
complement(48265..48550)
/rpt_family="Alu"
49630..50221
/rpt_family="MER41"
repeat_region
50781..51096
/rpt_family="Alu"
repeat_region
complement(51219..51439)
/rpt_family="Alu"
51462..51708
/rpt_family="MER33"
repeat_region
complement(52042..52321)
/rpt_family="Alu"
54001..54092
/rpt_family="Alu"
repeat_region
complement(55760..56014)
/rpt_family="Alu"
56616..56899
/rpt_family="Alu"
repeat_region
complement(66463..66747)
/rpt_family="Alu"
repeat_region
complement(66908..67180)
/rpt_family="Alu"

AF109119.1:62. .137,237. .2033)
/gene="HRNBPI"
/product="hexaribonucleotide binding protein 1 isoform
alpha"
/note="alternatively spliced"
join(AF109107.1:625. .711,AF109108.1:473. .415,
AF109109.1:166. .309,AF109110.1:94. .107,AF10911.1:75. .167,
AF109112.1:50. .110,AF109113.1:200. .253,AF109114.1:22. .102,
AF109115.1:67. .199,AF109116.1:108. .150,AF109117.1:35. .99,
AF109119.1:62. .137,159. .281)
/gene="HRNBPI"
/codon_start=1
/product="hexaribonucleotide binding protein 1 isoform
gamma"
/protein_id="AAL83409.1"
/db_xref="GI:19032386"
/translation="MLASQGVLLHPYGVPMIVPAAPVYGLIQQNQEAAPADTMAOP
VASAPFAPQNCIPAEYTHAPHAPAEYTGQTVVEHNLNYPRAQTHSESPADTSA
QVSGFATQDDAAPTDCGQPTQPSSENTENKSQPRRLIVSNIPFRFPDRLRMFGQF
GKILVEEITFNERGSKGFGFTFENSADADRAEKLHGTVEGKRIEYNNATYAMVN
KKTVAPLYNGKMLNPVGAUVSPFVYAGTALLCQANQSSSYSPSSLYTSAWPGF
PVPATATAAAYGALHRRGGRVTVTFPAAAPRPPIEYGGVVVQVEPYGKKLLGGY
AAVRYAQPATPATAAAYSDSYGKRVVADYHHALAPAEYGVGAMASITRGYNRFAPY
ADGVLSITLSDASITRGYNRFAPY"
join(AF109107.1:625. .711,AF109108.1:473. .415,
AF109109.1:166. .309,AF109110.1:94. .107,AF10911.1:75. .167,
AF109112.1:50. .110,AF109113.1:200. .253,AF109114.1:22. .102,
AF109115.1:67. .199,AF109116.1:108. .150,AF109117.1:35. .99,
AF109119.1:62. .137,237. .281)
/gene="HRNBPI"
/note="similar to Caenorhabditis elegans fox-1;
alternatively spliced"
/codon_start=1
/product="hexaribonucleotide binding protein 1 isoform
alpha"
/protein_id="AAL83408.1"
/db_xref="GI:19032385"
/translation="MLASQGVLLHPYGVPMIVPAAPVYGLIQQNQEAAPADTMAOP
VASAPFAPQNCIPAEYTHAPHAPAEYTGQTVVEHNLNYPRAQTHSESPADTSA
QVSGFATQDDAAPTDCGQPTQPSSENTENKSQPRRLIVSNIPFRFPDRLRMFGQF
GKILVEEITFNERGSKGFGFTFENSADADRAEKLHGTVEGKRIEYNNATYAMVN
KKTVAPLYNGKMLNPVGAUVSPFVYAGTALLCQANQSSSYSPSSLYTSAWPGF
PVPATATAAAYGALHRRGGRVTVTFPAAAPRPPIEYGGVVVQVEPYGKKLLGGY
AAVRYAQPATPATAAAYSDSYGKRVVADYHHALAPAEYGVGAMASITRGYNRFAPY
ADGVLSITLSDASITRGYNRFAPY"
join(AF109107.1:625. .711,AF109108.1:473. .415,
AF109109.1:166. .309,AF109110.1:94. .107,AF10911.1:75. .167,
AF109112.1:50. .110,AF109113.1:200. .253,AF109114.1:22. .102,
AF109115.1:67. .199,AF109116.1:108. .150,AF109117.1:35. .99,
AF109118.1:124. .176,AF109119.1:62. .137,159)
/gene="HRNBPI"
/note="similar to Caenorhabditis elegans fox-1;
alternatively spliced"
/codon_start=1
/product="hexaribonucleotide binding protein 1 isoform
beta"
/protein_id="AAL83407.1"
/db_xref="GI:19032384"
/translation="MLASQGVLLHPYGVPMIVPAAPVYGLIQQNQEAAPADTMAOP
VASAPFAPQNCIPAEYTHAPHAPAEYTGQTVVEHNLNYPRAQTHSESPADTSA
QVSGFATQDDAAPTDCGQPTQPSSENTENKSQPRRLIVSNIPFRFPDRLRMFGQF
GKILVEEITFNERGSKGFGFTFENSADADRAEKLHGTVEGKRIEYNNATYAMVN
KKTVAPLYNGKMLNPVGAUVSPFVYAGTALLCQANQSSSYSPSSLYTSAWPGF
PVPATATAAAYGALHRRGGRVTVTFPAAAPRPPIEYGGVVVQVEPYGKKLLGGY
AAVRYAQPATPATAAAYSDSYGKRVVADYHHALAPAEYGVGAMASITRGYNRFAPY
ADGVLSITLSDASITRGYNRFAPY"
join(AF109107.1:625. .711,AF109108.1:473. .415,
AF109109.1:166. .309,AF109110.1:94. .107,AF10911.1:75. .167,
AF109112.1:50. .110,AF109113.1:200. .253,AF109114.1:22. .102,
AF109115.1:67. .199,AF109116.1:108. .150,AF109117.1:35. .99,
AF109118.1:124. .176,AF109119.1:62. .137,159)
/gene="HRNBPI"
/note="similar to Caenorhabditis elegans fox-1;
alternatively spliced"
/codon_start=1
/product="hexaribonucleotide binding protein 1 isoform
beta"
/protein_id="AAL83407.1"
/db_xref="GI:19032384"
/translation="MLASQGVLLHPYGVPMIVPAAPVYGLIQQNQEAAPADTMAOP
VASAPFAPQNCIPAEYTHAPHAPAEYTGQTVVEHNLNYPRAQTHSESPADTSA
QVSGFATQDDAAPTDCGQPTQPSSENTENKSQPRRLIVSNIPFRFPDRLRMFGQF
GKILVEEITFNERGSKGFGFTFENSADADRAEKLHGTVEGKRIEYNNATYAMVN
KKTVAPLYNGKMLNPVGAUVSPFVYAGTALLCQANQSSSYSPSSLYTSAWPGF
PVPATATAAAYGALHRRGGRVTVTFPAAAPRPPIEYGGVVVQVEPYGKKLLGGY
AAVRYAQPATPATAAAYSDSYGKRVVADYHHALAPAEYGVGAMASITRGYNRFAPY
ADGVLSITLSDASITRGYNRFAPY"
159. .2033
/gene="HRNBPI"
/note="alternate exon 14"
/number=14
237. .2033
/gene="HRNBPI"
/note="alternate exon 14"

polyA_signal	/number=14	2017	.2022
polyA_site	/gene="HRNBp1"	2033	
BASE COUNT	922 a	521 C	586 G 942 T
ORIGIN			
Query Match	11.5%;	Score 70;	DB 9; Length 2971;
Beet Local Similarity	100.0%;	Pred. No. 1,5e-29;	
Matches 70;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
Oy	511	GATGCAAGACTAGAGGCATGTCATGATGTGGATCGTTCCTTCCTTCATTCAGAGCT	570
Db	180	GATGCCAAGACTAGAGGCATGTCATGATGTGGATCGTTCCTTCCTTCATTCAGAGCT	239
Oy	571	AGTATATATACC	580
Db	240	AGTATATATACC	249
RESULT 13			
LOCUS	AF109106	3348 bp	mRNA linear PRI 01-MAR-2002
DEFINITION	Homo sapiens hexaribonucleotide binding protein 1 isoform beta		
ACCESSION	AF109106		
VERSION	AF109106		
KEYWORDS	AF109106.1	GI:19032367	
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens.		
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Chen, W., Chu, Z., -L., Blough, R. I., Liu, L., Hoppes, B. and Winkelmann, J. C.		
TITLE	Molecular Cloning and Chromosomal Localization of a Novel Human Brain, Heart and Skeletal Muscle Specific RNA Binding Protein Gene Homologous to fox-1 in Caenorhabditis elegans		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 3348)		
AUTHORS	Chen, W., Chu, Z., -L., Blough, R. I., Liu, L., Hoppes, B. and Winkelmann, J. C.		
TITLE	Direct Submission		
JOURNAL	Submitted (24-NOV-1998) Internal Medicine/Hematology-Oncology, University of Cincinnati College of Medicine, 231 Bethesda Ave. 0508, Cincinnati, OH 45267-0508, USA		
FEATURES	Location/Qualifiers		
source	1. .3348		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="16"		
	/map="16p13.3"		
	1. .3348		
	/gene="HRNBp1"		
gene	234. .1441		
	/gene="HRNBp1"		
	/note="similar to Caenorhabditis elegans fox-1"		
	/codon_start=1		
	/product="hexaribonucleotide binding protein 1 isoform beta"		
	/protein_id="AAU83406.1"		
	/db_xref="GI:19032368"		
	/translat="MLASQGVLLHPYGVPMIVPAAYLPGLIGNOEAAAAPDTMAOP		
	VAAQAPAPONGI PAETAPVHPAPPEYGTQTPPHLTNLVPAQTHSOSPADTSA		
	GTSGATGATPDAAAPDGGOPOTOPSENTENSGPKRLHVSNI PFRFRDPLRMFGOF		
	GKLIDVEITIDNNGSKRGFGVTEFENSADADRAEKHGYVERBKIVNNATARVNTN		
	KKIVNPTNGWKLNPVGAAYSPFYAGTVLLCQAQOEGSSMSAPSLSVYTSAMPEF		
	PYPAAPAAAAYRGARLGRGRITVNTFRAAAPPP1PAYGCVGYDEPVGKMLLOGGY		
	AAARVYQPTPATAAYSDBRNQFVFAADEISCTSAVTDEFMLEPTTTHLLOPPPTA		
	LTVN"		
	3299. .3304		
	/gene="HRNBp1"		
polyA_signal			

REFERENCE 2 (pages 1 to 2279)
AUTHORS Shibata, H., Huynh, D. P., Vo, T. T. and Puls, S.-M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1998) Division of Neurology and Neurogenetics

Medical Center, UCLA School of Medicine, 8700 Beverly Blvd., Los Angeles, CA 90048, USA

FEATURES	Location/Qualifiers
source	1. .2279

```

gene      /note="similar to Homo sapiens cDNA clone HSP38A20"
1. .2279
/feature="A2BP"
1047. .2180
CDS

```

```

/translation="MAQCPYASAOFPAPDONGAIPAEYTAHPHPAEYAGQCTVPEPHTLNL
LYPPQTSISEGSPADTSAGTUSGTAIPDNGAPDGOPTOISENTEKRSQPKRHLAS
NIPFFRDPDLRQMFQDGLIDBAILTENEGSKQFEGYTTENSADADRAREKHLGTVY
VEGRILVENNMATAPVMTNKVTNPTAATYTNWKNMLPVGAAYSPEFAGVATLQANOSEDS
SMYASPSLVTYSARQEPYPATATPAAVYLRGHLRGRVYTYNFRVLAAPPQPI.PAGG
GVVYDDGFGGADIVGGVAAYRAOCTPPTAAVYASVSYRVAAYADPYHHALAPATYGGI
GAMNAPFLTDPAKTSHADVYGLVLSLQASLYRCGVNRFAPY"
nisc.feature
1341..1358

```

```
misc_feature      1452.  .1472      /note="Region: RNA binding motif RNP-2"
```

```

/Note="Region: RNA binding motif RNP-1"
BASE COUNT      497 a      712 c      636 g      434 t
ORIGIN

```

Query Match	11.5%	Score 70	DB 9	Length 2279
Best Local Similarity	100.0%	Pred. No. 1.5e-29		
Matches 70	Conservative 0	Mismatches 0	Indels 0	Gaps 0

Oy 511 GATGCCAAGACTAGAGCCCATCTGATGTGGGTCTCGTTCTTTCATTGCAGGCT 570
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2079 GATGCCAAGACTAGAGCCCATCTGATGTGGGTCTCGTTCTTTCATTGCAGGCT 2138

QY	571	AGTATATACC	580
Db	2139	AGTATATACC	2148

RESULT 11			
ARI34676			
LOCUS	ARI34676	2372 bp	DNA
DEFINITION	Sequence 1 from patent US 6194171.		
			linear
			PAT 16-MAY-2001

VERSION	ARI34676.1	GI:14123581
KEYWORDS		
SOURCE	Unknown.	
ORGANISM	Unknown.	

REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES
1 (bases 1 to 2372)	Pulst,S.M. and Shibata,H.	Nucleic acids encoding ataxin-2 binding proteins	Patent: US 6194171-A 1 27-FEB-2001;	Location/Qualifiers

BASE COUNT	548 a	726 c	646 g
ORIGIN	/organism="unknown"		

Query Match	11.5%;	Score 70;	DB 6;	Length 2372;
Best Local Similarity	100.0%;	Pred. No. 1.5e-29;		
Matches	70;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0

Qy	511	GATCCAAAGCTGAGAGCCATGCTGATATATGAGGGCTGGTCTTTCTTCATTCGAGGCT	570
Db	2133	GATCCAAAGCTGAGAGCCATGCTGATATATGAGGGCTGGTCTTTCTTCATTCGAGGCT	2192
Qy	571	AGTATATACC	580
Db	2193 <td>AGTATATACC</td> <td>2202</td>	AGTATATACC	2202

	RESULT	12
LOCUS	HSFOX14	
DEFINITION	Homo sapiens hexaribonucleotide binding protein 1 alpha, beta, and gamma isoforms (unpublished)	
	2971 bp	DNA linear
	PRI 01-MAR-2002	

ACCESSION	AF109120	GI:19032383
VERSION	AF109120.1	
KEYWORDS		
SEGMENT		
SOURCE	14 of 14	
	Homo sapiens.	

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE
1 (bases 1 to 2971)	Chen, W. and Winkelman, J. C.	The Exon-Intron Organization of the Human HRNBp1 Gene	Unpublished	2 (bases 1 to 2971)

TITLE
Molecular Cloning and Chromosomal Localization of a Novel Human Brain, Heart and Skeletal Muscle Specific RNA Binding Protein Gene Homologous to fox-1 in *Caenorhabditis elegans*

JOURNAL
Unpublished

DATE
2/15/2011

REVISION
2/15/2011

AUTHORS Chen, W. and Winkelmann, J. C.
TITLE Direct Submission
JOURNAL Submitted (25-NOV-1998) Internal Medicine/Hematology-Oncology,
University of Cincinnati College of Medicine, 231 Bethesda Ave., ML
0508, Cincinnati, OH 45267-0508, USA
Location/Analysis

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p13.3"
order:AF109107 1.372 948 AF109108 1:1 470.

```

```

AF109115.1:.. 251,AF109116.1:1.. 249,AF109117.1:1.. 151,
AF109118.1:1.. 221,AF109119.1:1.. 224,1.. 2033)
/gene="HRNBP1"
join(AF109107.1:372.. 711,AF109108.1:173.. 415,
AF109109.1:166.. 309,AF109110.1:94.. 147,AF109111.1:75.. 166,

```

AF1091115.1:1:67. .129,AF1091116.1:108. .130,AF1091117.1:55. .131,AF1091118.1:124. .176,AF1091119.1:62. .137,159. .2033)
/gene="HRNP1"
/product="hexaribonucleotide binding protein 1 isoform beta"

mRNA
 JOIN,AF1091107.1:1:372. /11,AF1073408.1:1:743. /12
 AF1091109.1:1:166. /309,AF1101810.1:1:194. /147,AF109111.1:1:75. /16
 AF109112.1:1:50. /10,AF109113.1:1:200. /251,AF109114.1:1:22. /10
 AF109115.1:1:67. /199,AF109116.1:1:108. /150,AF109117.1:1:35. /99
 AF109119.1:1:62. /137,159. /2033)

MRNA	Gravid
Join(AFI09107.1:372..	711, AFI09108.1:173.. 415,
AFI09109.1:166.. 309	AFI09110.1:94.. 147, AFI09111.. 1:75.. 16
AFI09112.1:50.. 110	AFI09113.1:200.. 255, AFI09114.1:22.. 10
AFI09115.1:67.. 199	AFI09116.1:108.. 150, AFI09117.1:35.. 99

LOCUS AK001027 1513 bp mRNA linear PRI 01-AUG-2002
DEFINITION Homo sapiens cDNA FLJ10165 f18, clone HEMBA1003591, weakly similar
to CHLOROPLAST 28 KD RIBONUCLEOPROTEIN PRECURSOR.
ACCESSION AK001027
VERSION AK001027.1 GI:7022045
KEYWORDS oligo capping; f18 (full insert sequence).
SOURCE Homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to
mRNA, clone_11b:HEMBA1 clone:HEMBA1003591.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Iisagai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H.,
Wagatsuma, M., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Chiba, Y., Ichida, S., Murakami, K.,
Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K.,
Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A.,
Nakamura, Y., Nagahara, K., Masuno, Y., Ninomiya, K. and Iwayanagi, T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 1513)
Iisagai, T. and Otsuki, T.
Direct Submission
Submitted (16-FEB-2000) Takao Iisagai, Helix Research Institute,
Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: cDNA full insert
sequencing: Research Association for Biotechnology, cDNA library
construction, 5' - 3' end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
source
location/Qualifiers
1. .1513
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1003591"
/cissue="whole embryo, mainly head"
/clone_lib="HEMBA1"
/dev_stage="embryo, 10 weeks"
/note="cloning vector: pME185FL3"
21. .141
/note="unnamed protein product"
/codon_start=1
/protein_id="BAA91472.1"
/db_xref="GI:7022046"
/translation="MNCERQLRGNQEAAPPTMAQPVASQFAPPPONGIPAEVTAP
HPHPABEYTGQTVPEHTLNLPPAQTSHSOSPADTSAQVSGTATQADDAAPTDGQ
OTQPSNTENKSKOPKRLHVNIPERFDPDLROMFGQIGLIVEIIFNRSKGFQF
VTFENSADNRAREKLGHTVEGRKIEVNNATKTVNPKTVNPGMKNLVNCGV
XSPREYAGTVLLCOANOQSSWMSABSLVYTGAMGCPRPATATAAAYGALRBC
RTVNTFRAPAPPPPIPAVGIVYDDFVADYGGGATRIYAQPPATATAASDSIG
RVYAADPPHHLAPAPYGVGAMNAFAPLTDATRSHADVGLVLSLQASIVRGYN
RFAPY"
BASE COUNT 354 a 418 c 395 g 346 t
ORIGIN
Query Match 11.5%; Score 70; DB 9; Length 1513;
Best Local Similarity 100.0%; Pred. No. 1.5e-29;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

AF229057
LOCUS AF229057 2000 bp mRNA linear PRI 01-FEB-2002
DEFINITION Homo sapiens hexaribonucleotide binding protein 1 isoform gamma
(HRNBP1) mRNA, complete cds.
ACCESSION AF229057
VERSION AF229057.1 GI:18461366
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2000)
AUTHORS Chen, W. and Winkelman, J. C.
Direct Submission
Submitted (28-JAN-2000) Internal Medicine/Hematology-Oncology,
University of Cincinnati College of Medicine, The Vontz Center for
Molecular Studies, 3125 Eden Avenue, Cincinnati, OH 45267-0508, USA
FEATURES
source
location/Qualifiers
1. .2000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p13.3"
1. .2000
/gene="HRNBP1"
256. .1512
/gene="HRNBP1"
/note="RNA-binding protein"
/codon_start=1
/product="hexaribonucleotide binding protein 1 isoform
gamma"
/protein_id="AAU71904.1"
/db_xref="GI:18461367"
/translation="MLASQGVLLHPYGVPMIVPAAYLPLGLQGNQEAAPPTMAQ
VYSAQFAPPPONGIPAEVTAPHPHPABEYTGQTVPEHTLNLPPAQTSHSOSPADTSA
QVSGTATQADDAAPTDGQOTQPSNTENKSKOPKRLHVNIPERFDPDLROMFGQ
GILIDVELIFNERSKGFVTFENSADNRAREKLGHTVEGRKIEVNNATKTVN
KTVNPTNGMKNLPPVGAIVYSEFVAGTVLLCOANOQSSWMSABSLVYTGAMG
PYPAATAAAYGALRBCRTVNTFRAPAPPPPIPAVGIVYDDFVADYGGGATRIYAQPPATATAASDSIG
AAVRYAADPPHHLAPAPYGVGAMNAFAPLTDATRSHADVGLVLSLQASIVRGYNRFAPY"
BASE COUNT 502 a 505 c 487 g 506 t
ORIGIN
Query Match 11.5%; Score 70; DB 9; Length 2000;
Best Local Similarity 100.0%; Pred. No. 1.5e-29;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 511 GATGCCAAGACTAGAGCCATGCTGATGATGGGCTCGTTCTTTCTTATTCAGAGCT 570
DB 1411 GATGCCAAGACTAGAGCCATGCTGATGATGGGCTCGTTCTTTCTTATTCAGAGCT 1470
QY 571 AGTATATACC 580
DB 1471 AGTATATACC 1480

RESULT 10
AF107203 2279 bp mRNA linear PRI 23-JUN-2000
LOCUS AF107203
DEFINITION Homo sapiens ataxin-2-binding protein (A2BP) mRNA, complete cds.
ACCESSION AF107203
VERSION AF107203.1 GI:8671585
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2279)
AUTHORS Shibata, H., Huynh, D. P., Vo, T. T. and Puls, S.-M.
TITLE A novel protein, A2BP, with RNA binding motif binds to C-terminal
ataxin-2
JOURNAL Unpublished

* 32686 34092: contig of 1407 bp in length
* 34093 34192: gap of unknown length
* 34193 34491: contig of 1299 bp in length
* 35492 35591: gap of unknown length
* 35592 37096: contig of 1505 bp in length
* 37097 37196: gap of unknown length
* 37197 38312: contig of 1116 bp in length
* 38313 40579: gap of unknown length
* 38413 40579: contig of 2167 bp in length
* 40580 40679: gap of unknown length
* 40680 42711: contig of 2032 bp in length
* 42712 42811: gap of unknown length
* 42812 44224: contig of 1413 bp in length
* 44225 44324: gap of unknown length
* 44325 45859: contig of 1535 bp in length
* 45860 45959: gap of unknown length
* 45960 47975: contig of 2016 bp in length
* 47976 48075: gap of unknown length
* 48076 50049: contig of 1974 bp in length
* 50050 50149: gap of unknown length
* 50150 51700: contig of 1551 bp in length
* 51701 51800: gap of unknown length
* 51801 53551: contig of 1751 bp in length
* 53552 53651: gap of unknown length
* 53652 56127: contig of 2476 bp in length
* 56128 56227: gap of unknown length
* 56228 57920: contig of 1693 bp in length
* 57921 58020: gap of unknown length
* 58021 60257: contig of 2237 bp in length
* 60258 60357: gap of unknown length
* 60358 62838: contig of 2481 bp in length
* 62839 62938: gap of unknown length
* 62939 64839: contig of 1901 bp in length
* 64840 64939: gap of unknown length
* 64940 66877: contig of 1938 bp in length
* 66878 66978: gap of unknown length
* 66979 69856: contig of 2879 bp in length
* 69857 71716: gap of unknown length
* 71717 71816: contig of 1760 bp in length
* 71817 73929: gap of unknown length
* 73929 73929: contig of 2113 bp in length.
Location/Qualifiers
1..73929
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-33J14"
BASE COUNT 19038 a 16649 c 14558 g 19363 t 4321 others
ORIGIN
Query Match 13.0% Score 79; DB 2; Length 73923;
Best Local Similarity 100.0%; Pred. No. 5.9e-35;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 411 CAGTTACGACGAGATTATGTCGCGACCCCTTACACACACACACTGCTCCAGCCCCCAG 470
DB 11590 CAGTTACGACGAGATTATGTCGCGACCCCTTACACACACACACTGCTCCAGCCCCCAG 11649
QY 471 CTACGCGCTTGTCGCATG 489
DB 11650 CTACGCGCTTGTCGCATG 11668
RESULT 7
LOCUS AB060859 1475 bp mRNA linear PRI 13-JUN-2001
DEFINITION Macaca fascicularis brain cDNA clone:Qlra-11594, full insert
sequence.
ACCESSION AB060859
VERSION AB060859.1 GI:13874510
KEYWORDS oligo capping; fls (full insert sequence).
SOURCE Macaca fascicularis adult male temporal lobe right cDNA to mRNA,
clone_lib:macaque brain cDNA library Qlra clone:Qlra-11594.
ORGANISM Macaca fascicularis

REFERENCE
AUTHORS Osada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirai,M.,
Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.
TITLE Isolation of full-length cDNA clones from macaque brain cDNA
libraries
JOURNAL Unpublished
AUTHORS Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.
TITLE Direct Submission
JOURNAL Submitted (27-APR-2001) Katsuyuki Hashimoto, National Institute of
Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
1-chome, Shinyuku-ku, Tokyo 162-8640, Japan
(E-mail:hashi@nih.go.jp, URL:http://www.nih.go.jp/yoken/genbank/
Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)
Lab host: TOP10
Vector: pME185-FL3 (Acc.No. AB009864)
R. Site1: DraIII (CACTGTGTG)
R. Site2: DraIII (CAGCATGTG)
Description: 1st strand cDNA was primed with an oligo(dt) primer
[ATGCGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was synthesized
using specific 5' and 3' primers and amplified by PCR. The PCR
product was digested with SfiI and size selection was performed to
exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
into distinct DraIII sites of pME185-FL3. XhoI sites just outside
the DraIII sites can be used to isolate the cDNA insert. Libraries
were constructed by oligo-capping method
(Sugano et al., Institute of Medical Science, University of
Tokyo).
Custom primer used for sequencing
(5' end primer [CTTCTGCTCTAAAGCTGCG];
3' end primer [CGATCGACGCTGACGACA]).
Location/Qualifiers
1..1475
/organism="Macaca fascicularis"
/db_xref="taxon:9541"
/clone="Qlra-11594"
/sex="male"
/tissue-type="temporal lobe right"
/clone_lib="macaque brain cDNA library Qlra"
/dev_stage="adult"
90..1220
/codon_start=1
/product="hypothetical protein"
/protein_id="BAB46877.1"
/db_xref="GI:13874511"
/translation="MEERGRKRVQGGNOEAAADPTMAQPYASNOFAPPNQICPAEYT
APHPAPAEYGTQTPVEHTLNLPAPQTHSESGPADTNAQTVSGTATQDDAAPTDS
QPTQPSBENTKQPRKLHVSNIPEFRPDLDROMGOFGKILDVELIENERSKGF
GFVPEFNSADADRAREKLHGVGRKLEVNATARVNTKNTAPYTNKWLNVVVG
AVVSEPEFYAGTVLLICANOGESSWYSPSSLVTSAMGPFYPATATAAAYRGALRG
RGRTVNTFRRAAPPPPIPRATGVTYODGDFGADIGCYGAAYRAQTPATPAAYSDR
NOFFVVADEISCTVSAVTDFEFLPPTTHLLOPPPALVP"
BASE COUNT 389 a 414 c 363 g 309 t
ORIGIN
Query Match 11.5% Score 70; DB 9; Length 1475;
Best Local Similarity 100.0%; Pred. No. 1.5e-29;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 511 GATCCCAAGACTAGAGCATGCTGATGATGCGCTCTCTTCTTCATTCAGAGCT 570
DB 1241 GATCCCAAGACTAGAGCATGCTGATGATGCGCTCTCTTCTTCATTCAGAGCT 1300
QY 571 AGTATATACC 580
DB 1301 AGTATATACC 1310
RESULT 8
AK001027

REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
AUTHORS	Oseada, N., Kueuda, J., Tanuma, R., Ito, A., Hirata, M., Sugano, S. and Hashimoto, K.
TITLE	Isolation of full-length cDNA clones from mouse brain cDNA library made by oligo-capping method
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 2002)
AUTHORS	Hashimoto, K., Oseada, N., Kueuda, J. and Sugano, S.
TITLE	Direct Submission
JOURNAL	Submitted (12-Apr-2000) Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan (E-mail: khashi@nih.go.jp URL: http://www.nih.go.jp/yoken/genebank/, Tel.: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181) URL: http://www.nih.go.jp/yoken/genebank/
COMMENT	Lab Name: Sugano mouse brain mncb Lab host: TOPI0 Vector: pME18S-FLJ 1st strand cDNA was primed with an oligo(dT) primer [ATGTCGCCCTTTTGTTTTTTTTTT]; double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18S-FLJ. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Library was constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing (5' end primer [CTTGCGCTCTAAAGCTGGC] ; 3' end primer [CGACCTGCAGCTGCAGCAC]). A part of this sequence is reported in AU067167.
FEATURES	Location/Qualifiers 1..2002 /organism="Mus musculus" /strain="C57BL" /db_xref="taxon:10090" /clone="MNcb-3035" /sex="female" /clone_1lb="Sugano mouse brain mncb" /dev_stage="adult" 637..1827 /note="unnamed protein product" /codon_start=1 /protein_id="BA95079.1" /db_xref="GI:7670456" /translation="MNCERQLRQNEAAMPDTMAOPYASQAPFAPONGIPAEYTAARHPHAPRYGQTIVGHILANTLPPTDISEBSADTSAGTVSGATQTDADAIPDGQCTOPENSTESKQPKRLHNSNPFRPDIDLOMGOFGLIDVELIFNERSGKGFGVAF TFEASADVADRAREKLHGTVNGRKIEVNVAATMTKNKTANPYTMGNKINPVGAYVY SPDYAGATVLICQANOESSNTPSSSLVTYSAMRPFPYPATTAAYVRGMHLRGGRGV TVNVTFRRAAPPRPIPAYGAVYQDDGFAGAIYGGYAAYRAOFTPATLAAYSISYGVYVADPPHHITLAPAITYGVAMNAFAPLTDAKTRGHADVGLVLSLQASLYRGGYNRFAPY"
CDS	
BASE COUNT	518 a 589 c 516 g 379 t
ORIGIN	
Query Match	31.9%; Score 194; DB 10; Length 2002;
Best Local Similarity	99.6%; Pred. No. 3e-103;
Matches 244; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
Db	22 AACCCCTACCAATGCGTCGGAATAATTAAATCCAGTTGTGGGGCGGCTACAGCCCCGAC 81
Db	1240 AACCCCTACCAATGCGTCGGAATAATTAAATCCAGTTGTGGGGCGGCTACAGCCCCGAC 1299
Db	82 TTCATGACGACGAGCGTGTGTGTGGCCAGGCAACAGAGAGGATCTCCATGTACAGT 141
Db	1300 TTCTATGACGACGACGCTGTGTGTGGCCAGGCAACAGAGAGGATCTTCATGTACAGT 1359
Db	142 GGCCCGAGTTCACTGTATATATCTTGCAGTACCCTGGCTTTCATATCCGCGCGCACT 201
Db	1360 GGCCCGAGTTCACTGTATATATCTTGCAGTACCCTGGCTTTCATATCCGCGCGCACT 1419
Db	202 GCTGCAGCTGCATACCGAGGCGCTACCTTCGAGCGCGTGTGCCAGCCGTGTACAACACC 261

Db	1420	GCTGAGCTGATACCCAGAGGGGCTCACCCTTGAGAGCGGTGTGCGACCGTGTACAAACACC	1479
Oy	262	TTCAG 266	
Db	1480	TTCAG 1484	
RESULT 4			
LOCUS	AF191501	1586 bp	linear ROD 01-MAR-2002
DEFINITION	Mus musculus hexaribonucleotide binding protein 1 (Hrbp1) mRNA,		
ACCESSION	AF191501		
VERSION	AF191501.1	GI:19032413	
KEYWORDS			
SOURCE	Mus musculus.		
ORGANISM	Mus musculus		
REFERENCE	Chen, W. and Winklermann, J. C.		
AUTHORS	Submitted (01-OCT-1999) Internal Medicine/Hematology-Oncology,		
TITLE	University of Cincinnati College of Medicine, 231 Bethesda Ave.,		
JOURNAL	Cincinnati, OH 45267-0509, USA		
FEATURES			
source	Location/Qualifiers		
	1..1586		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/ribase_type="skeletal muscle"		
	1..1586		
	/gene="Hrbp1"		
	166..1419		
	/gene="Hrbp1"		
	/codon_start=1		
	/product="hexaribonucleotide binding protein 1"		
	/protein_id="AA183425.1"		
	/db_xref="GI:19032414"		
	/translation="MLASQGVLLHSYGVPMIVPAAPYFPGIMQGNQEAAPDTPAQP		
	YASQAPDPNGIPEAYTAPHPAPRYTQTTVPDHTLNLPEPTQHSBSADTSQ		
	TVSGATQITDDAAPDTGDPOTPOPTSEMSKSPKRLHVSNIPEFRDPDIRMGCGFG		
	KILDELITENERSKRGFGVTFENSADADAREKLTHTVEGKRIEYNNATATVMYTKR		
	KTNPTPTNGKMLPNVGAAYVSPDYACTVLICANQBSMYSGPSLVTYSAMPGR		
	YAPASAAAYRGAHLKRGRTGNTVLAHPPTPIFAYGAVVIOEPYRKLKLOGGTA		
	AYRYAOPTPATAAAYSDSYGVRYAADYHHTLAPPIYGVGMNAFAPLTDAKTRSHA		
	DDVGLVLSISQASIVRGVNRFPAPY"		
BASE COUNT	407 a 458 c 374 g 347 t		
ORIGIN			
Query Match	26.4%; Score 161; DB 10; Length 1586;		
Best Local Similarity	100.0%; Pzed. No. 1.2e-83;		
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Oy	22	AACCCCTACACCAATGGGTGAATTAATCACTGCTGGGGCGGCTACAGCCCGCAG	81
Db	829	AACCCCTACACCAATGGGTGAATTAATCACTGCTGGGGCGGCTACAGCCCGCAG	888
Oy	82	TTCTATGAGGACGAGGCTGTTGTGCGAGGCCAACACGAGAGGATCTTCATGTACAGT	141
Db	889	TTCTATGAGGACGAGGCTGTTGTGCGAGGCCAACACGAGAGGATCTTCATGTACAGT	948
Oy	142	GGCCCACTTCACTTGTATATATCTTTCGCAATGCGTGGCTT	182
Db	949	GGCCCACTTCACTTGTATATATCTTTCGCAATGCGTGGCTT	989
RESULT 5			
LOCUS	AC120661/c	115860 bp	DNA linear HTG 23-JUL-2002
DEFINITION	Rattus norvegicus clone CH230-34C2, *** SEQUENCING IN PROGRESS ***,		
ACCESSION	AC120661		

JOURNAL Patent: WO 0174901-A 1 11-OCT-2001;
Scios Inc. (US)
Location/Qualifiers
FEATURES 1..1340
source /organism="Rattus norvegicus"
/db_xref="taxon:10116"
BASE COUNT 366 a 394 c 317 g 263 t
ORIGIN

Query Match 100.0%; Score 609; DB 6; Length 1340;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACTAATAAAGGCGCTGAACCCCTACCAATGCTGGAATTAATCAAGTTGTG 60
DB 535 ATGACTAATAAAGGCGCTGAACCCCTACCAATGCTGGAATTAATCAAGTTGTG 594
QY 61 GCGCGGCTCTACAGCCCGCACTTTATGACAGCAGCGTCTGTTGCGAGGCAACAG 120
DB 595 GCGCGGCTCTACAGCCCGCACTTTATGACAGCAGCGTCTGTTGCGAGGCAACAG 654
QY 121 GAGGATCTTCATGATGACGTGGCCCACTTGAATTAATCAAGTTGTG 180
DB 655 GAGGATCTTCATGATGACGTGGCCCACTTGAATTAATCAAGTTGTG 714
QY 181 TTTCATATCCGCGCGCACTGCTGAGCTGATACCGAGGCGCTCACTTGAAGCCGT 240
DB 715 TTTCATATCCGCGCGCACTGCTGAGCTGATACCGAGGCGCTCACTTGAAGCCGT 774
QY 241 GGTGCGACCGTGTACACACCTTTCAGAGCTGGCGGCGCCCAATCCCGGCTAT 300
DB 775 GGTGCGACCGTGTACACACCTTTCAGAGCTGGCGGCGCCCAATCCCGGCTAT 834
QY 301 GCGGAGTGTGTATGAAGCCAGGTATGGAATTAATGCTAGAGGTGTTACGCT 360
DB 835 GCGGAGTGTGTATGAAGCCAGGTATGGAATTAATGCTAGAGGTGTTACGCT 894
QY 361 GCATACCGCTAGCGCCAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
DB 895 GCATACCGCTAGCGCCAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 954
QY 421 CGAGTTATGCTGCGAGCCCTTACACACACCTTGTCCAGCCCACTTGAAGCCGT 480
DB 955 CGAGTTATGCTGCGAGCCCTTACACACACCTTGTCCAGCCCACTTGAAGCCGT 1014
QY 481 GGTGCGATGAAGCTTTTGGCCCTTACCGAAGCTAGAGCCATGCTGATGAT 540
DB 1015 GGTGCGATGAAGCTTTTGGCCCTTACCGAAGCTAGAGCCATGCTGATGAT 1074
QY 541 GTGGGCTGCTTCTTCTCATGAGGCTAGATATACCAAGGGGATACACCGTTT 600
DB 1075 GTGGGCTGCTTCTTCTCATGAGGCTAGATATACCAAGGGGATACACCGTTT 1134
QY 601 GCTCCATAT 609
DB 1135 GCTCCATAT 1143

RESULT 2
AF107204 1363 bp mRNA linear ROD 23-JUN-2000
LOCUS AF107204
DEFINITION Mus musculus ataxin 2-binding protein (A2bp) mRNA, complete cds.
ACCESSION AF107204
VERSION AF107204.1 GI:8671587
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Shibata,H., Huynh,D.P., Vo,T.T. and Pulst,S.-M.
TITLE A novel protein, A2BP, with RNA binding motif, binds to C-terminal ataxin-2

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1363)
AUTHORS Shibata,H., Huynh,D.P., Vo,T.T. and Pulst,S.-M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1998) Division of Neurology and Neurogenetics
Laboratory, Burns and Allen Research Institute, Cedars-Sinai
Medical Center, UCLA School of Medicine, 8700 Beverly Blvd., Los
Angeles, CA 90048, USA

FEATURES 1..1363
source Location/Qualifiers
1..1363
/organism="Mus musculus"
/db_xref="taxon:10090"
1..1363
gene /gene="A2bp"
67..1203
/gene="A2bp"
CDS /codon_start=1
/product="ataxin 2-binding protein"
/protein_id="AA078292.1"
/db_xref="GI:8671588"
/translation="MAQPVASQAPAPPPONGIPAEYTAAPHAPAEYTGQTPVPHLNL
LYPTQTHSEASADTSATVSGTATOTDDAAPTQGOPTOPESEKTSKOPRLVSN
IPPRFRDPDLRMFGQFGKILDELVEINERSGKGFYFENSADAPRELHGT
VBSRKLEVNNAIRVNTNKTVPTNMGKLVNVAIVGAVSPDYAGTVLCLQNOEG
SSWYSGPSLVYTSAMPGFYPAATAAARGAHLRGRVNTVTFRAAPPPPIPAY
GVAMNAPAPLTDATKRSADVDGVLSSLQSIYRGVNRAPY"
358..375
/gene="A2bp"
/note="Region: RNA binding motif RNP-2"
misc_feature 475..495
/gene="A2bp"
/note="Region: RNA binding motif RNP-1"

BASE COUNT 362 a 396 c 324 g 281 t
ORIGIN

Query Match 31.9%; Score 194; DB 10; Length 1363;
Best Local Similarity 99.6%; Pred. No. 3e-103;
Matches 244; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 22 AACCCCTACACCAATGCTGGAATTAATCCAGTTGGCGCGCTCTACAGCCCGAC 81
DB 616 AACCCCTACACCAATGCTGGAATTAATCCAGTTGGCGCGCTCTACAGCCCGAC 675
QY 82 TTCTATGAGGACGAGTGTGTTGCGAGGCAACAGAGAGATCTTCATGTAAGT 141
DB 676 TTCTATGAGGACGAGTGTGTTGCGAGGCAACAGAGAGATCTTCATGTAAGT 735
QY 142 GCGCCAGTTCACTTGTATATCTTGCATGCTGCTGCTTCCATATCCGCGCCACT 201
DB 736 GCGCCAGTTCACTTGTATATCTTGCATGCTGCTGCTTCCATATCCGCGCCACT 795
QY 202 GCTGCACTGCTATCCGAGGGGCTCACTTGAAGCCGTGTGCGACCGTGTACACACC 261
DB 796 GCTGCACTGCTATCCGAGGGGCTCACTTGAAGCCGTGTGCGACCGTGTACACACC 855
QY 262 TTCAG 266
DB 856 TTCAG 860

RESULT 3
AB041596 2002 bp mRNA linear ROD 30-JUN-2000
LOCUS AB041596
DEFINITION Mus musculus brain cDNA, clone MNCD-3035, similar to Homo sapiens
CDNA FLJ10165, clone HEMBA1003591.
ACCESSION AB041596
VERSION AB041596.1 GI:7670455
KEYWORDS
SOURCE Mus musculus (strain:C57BL/6J adult female cDNA to mRNA,
clone:lib:Sugano mouse brain mncd clone:MNCD-3035).
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 14:22:28 ; Search time 1627.96 Seconds
(without alignments)
10887.019 Million cell updates/sec

Title: US-09-809-545a-1_COPY_535_1143
Perfect score: 609
Sequence: 1 atgactaataaaaaagccgt.....acaacgcttgcctccatcat 609

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:

- 1: gb_da:*
- 2: gb_hcg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pac:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_ats:*
- 12: gb_ey:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_da:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pac:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_ats:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_hcg_hum:*
- 31: em_hcg_inv:*
- 32: em_hcg_other:*
- 33: em_hcg_mus:*
- 34: em_hcg_pln:*
- 35: em_hcg_rtd:*
- 36: em_hcg_mam:*
- 37: em_hcg_vtc:*
- 38: em_ay:*
- 39: em_hugo_hum:*
- 40: em_hugo_mus:*
- 41: em_hugo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	609	100.0	1340	6 AX268800	AX268800 Sequence
2	194	31.9	1363	10 AF107204	AF107204 Mus muscu
3	194	31.9	2002	10 AB041596	AB041596 Mus muscu
4	161	26.4	1586	10 AF191501	AF191501 Mus muscu
5	92	15.1	115860	2 AC120661	AC120661 Rattus no
6	79	13.0	73929	2 AC103207	AC103207 Rattus no
7	70	11.5	1475	9 AB060859	AB060859 Macaca fa
8	70	11.5	1513	9 AK001027	AK001027 Homo sapi
9	70	11.5	2000	9 AF229057	AF229057 Homo sapi
10	70	11.5	2279	9 AF107203	AF107203 Homo sapi
11	70	11.5	2372	6 AR134676	AR134676 Sequence
12	70	11.5	2971	9 HSFOX14	AF109120 Homo sapi
13	70	11.5	3348	9 AF109106	AF109106 Homo sapi
14	70	11.5	164538	9 AC005774	AC005774 Homo sapi
15	67	11.0	115860	2 AC120661	AC120661 Rattus no
16	54	8.9	1547	9 AF094849	AF094849 Homo sapi
17	43	7.1	1374	9 HSFOX08	AF109114 Homo sapi
18	43	7.1	1623	9 HSM803066	AF171700 Homo sapi
19	40	6.6	234	9 HSFOX13	AF109119 Homo sapi
20	26	4.3	251	9 HSFOX09	AF109115 Homo sapi
21	26	4.3	596	9 HSA323688	AJ323688 Homo sapi
22	26	4.3	617	9 HSA335811	AJ335811 Homo sapi
23	26	4.3	653	9 HSA335616	AJ335616 Homo sapi
24	26	4.3	675	9 HSA334013	AJ334013 Homo sapi
25	23	3.8	151	9 HSFOX11	AF109117 Homo sapi
26	23	3.6	154471	9 AC022523	AC022523 Homo sapi
27	22	3.6	184497	2 AC099688	AC099688 Homo sapi
28	22	3.6	210713	2 AC024271	AC024271 Homo sapi
29	21	3.4	105589	2 AC092269	AC092269 Homo sapi
30	21	3.4	157613	9 AC018769	AC018769 Homo sapi
31	21	3.4	166847	9 AC008393	AC008393 Homo sapi
32	21	3.4	180520	2 AC113426	AC113426 Homo sapi
33	21	3.4	186872	9 AC074052	AC074052 Homo sapi
34	21	3.4	188046	9 AC124857	AC124857 Homo sapi
35	21	3.4	195068	9 AC012100	AC012100 Homo sapi
36	21	3.4	197000	2 AC124859	AC124859 Homo sapi
37	21	3.4	207751	2 AC112189	AC112189 Homo sapi
38	20	3.3	249	9 HSFOX10	AF109116 Homo sapi
39	20	3.3	4684	9 AK001678	AK001678 Homo sapi
40	20	3.3	6123	2 AF303002	AF303002 Homo sapi
41	20	3.3	91531	2 AC079846_3	Continuation (4 of
42	20	3.3	116539	2 AC132329	AC132329 Rattus no
43	20	3.3	118369	2 AC131018	AC131018 Rattus no
44	20	3.3	139242	10 AL590996	AL590996 Mouse DNA
45	20	3.3	140186	2 AC055808	AC055808 Homo sapi

ALIGNMENTS

RESULT 1
AX268800
LOCUS AX268800 1340 bp
DEFINITION Sequence 1 from Patent WO0174901.
ACCESSION AX268800
VERSION AX268800.1 GI:16541860
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1
AUTHORS Stanton, L.W. and White, R.T.
TITLE Secreted factors

THIS PAGE BLANK (USPTO)

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 454
OTHER INFORMATION: n = A,T,C or G
US-09-998-598-2571

Query Match 5.1%; Score 31; DB 10; Length 466;
Best Local Similarity 57.9%; Pred. No. 5.5;
Matches 55; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 166 TCTGCAATGCTGCTTCCATATCCGCGCCACTGCTGCAGCTGCATACCGAGGGCT 225
DB 397 TTGCAACAGGCTGGTGTGCACTTGCAGACGTAATGTGCGAGGTGCGCAGGTT 338
QY 226 CACCTTCGAGGCCGTGTGCGACCGGTGTAACAAC 260
DB 337 GACCTCCGTGGCGAGGTAGTGCCTTCCACGC 303

Search completed: March 15, 2003, 16:12:27
Job time : 564.057 secs

APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA 011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 7757
LENGTH: 927
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(927)
US-09-815-242-7757

Query Match
Best Local Similarity 5.1%; Score 31.2; DB 10; Length 927;
Matches 81; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

Qy 341 TGTACAGGAGTGTACGCTGCATACCGCCAGCCACCCCTGCACCTGCTG 400
Db 395 TGTCCAGCAAGTCTGCTGGCCGCGCAGCAGCGGCGACAGTGGCGTTG 336
Qy 401 CCTACAGTACAGTACGACGAGTTTATGCTGCCACCCCTACACACACTTGT 460
Db 335 GGCACATATGACATGCGCTTGGCTGCTGCTGACACGATGACGCGCACT 276
Qy 461 CAGCCCCCAGCTACGCGTGTGACCATGATCTTTGGCGCC 504
Db 275 GACTCCCCGCGCACCCGCTGACGCGCTGACATGCAATGCCCTC 232

RESULT 13
US-09-557-232-3/C
Sequence 3, Application US/09557232
Publication No. US20030040099A1
GENERAL INFORMATION:
APPLICANT: Frey, Terry
APPLICANT: Pougatchev, Konstantin
APPLICANT: Abernathy, Emily
TITLE OF INVENTION: Highly Infectious Rubella Virus Clones and Methods of Productio
FILE REFERENCE: 07362-0104/43439-221816
CURRENT APPLICATION NUMBER: US/09/557,232
CURRENT FILING DATE: 2000-04-24
PRIOR APPLICATION NUMBER: 08/459,041
PRIOR FILING DATE: 1995-06-02
PRIOR APPLICATION NUMBER: 08/093,453
PRIOR FILING DATE: 1993-07-19
PRIOR APPLICATION NUMBER: 07/722,334
PRIOR FILING DATE: 1991-06-28
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patent version 3.0
SEQ ID NO: 3
LENGTH: 2558
TYPE: DNA
ORGANISM: Rubella virus

US-09-557-232-3

Query Match
Best Local Similarity 5.1%; Score 31.2; DB 9; Length 2558;
Matches 51; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 168 TGCATGCTGCTTTCATATCCGGCCGACATGCTGCACCTGCATACCGAGGGCTCA 227
Db 587 TGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 528
Qy 228 CCTCGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 251
Db 527 CCTCGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 504

RESULT 14
US-09-557-232-1/C
Sequence 1, Application US/09557232
Publication No. US20030040099A1
GENERAL INFORMATION:
APPLICANT: Frey, Terry
APPLICANT: Pougatchev, Konstantin
APPLICANT: Abernathy, Emily
TITLE OF INVENTION: Highly Infectious Rubella Virus Clones and Methods of Productio
FILE REFERENCE: 07362-0104/43439-221816
CURRENT APPLICATION NUMBER: US/09/557,232
CURRENT FILING DATE: 2000-04-24
PRIOR APPLICATION NUMBER: 08/459,041
PRIOR FILING DATE: 1995-06-02
PRIOR APPLICATION NUMBER: 08/093,453
PRIOR FILING DATE: 1993-07-19
PRIOR APPLICATION NUMBER: 07/722,334
PRIOR FILING DATE: 1991-06-28
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patent version 3.0
SEQ ID NO: 1
LENGTH: 9759
TYPE: DNA
ORGANISM: synthetic construct - rubella
US-09-557-232-1

Query Match
Best Local Similarity 5.1%; Score 31.2; DB 9; Length 9759;
Matches 51; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 168 TGCATGCTGCTTTCATATCCGGCCGACATGCTGCACCTGCATACCGAGGGCTCA 227
Db 3389 TGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3330

Qy 228 CCTCGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 251
Db 3329 CCTCGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3306

RESULT 15
US-09-998-598-2571/C
Sequence 2571, Application US/09998598
Patent No. US20020150922A1
GENERAL INFORMATION:
APPLICANT: Stolk, John A.
APPLICANT: Xu, Jiangchun
APPLICANT: Chenault, Ruth A.
APPLICANT: Meagher, Madelein Joy
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.561
CURRENT APPLICATION NUMBER: US/09/998,598
CURRENT FILING DATE: 2001-11-16
NUMBER OF SEQ ID NOS: 2606
SOFTWARE: Corixa Invention Disclosure Database
SEQ ID NO: 2571
LENGTH: 466
TYPE: DNA

APPLICANT: OZAKI, AKIO
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCE: 249-125
 CURRENT APPLICATION NUMBER: US/09/738,626
 CURRENT FILING DATE: 2000-12-18
 PRIOR APPLICATION NUMBER: JP 99/377484
 PRIOR FILING DATE: 1999-12-16
 PRIOR APPLICATION NUMBER: JP 00/159162
 PRIOR FILING DATE: 2000-04-07
 PRIOR APPLICATION NUMBER: JP 00/280988
 PRIOR FILING DATE: 2000-08-03
 NUMBER OF SEQ ID NOS: 7059
 SOFTWARE: PatentIn ver. 3.0
 SEQ ID NO 3418
 LENGTH: 1195
 TYPE: DNA
 ORGANISM: Corynebacterium glutamicum
 US-09-738-626-3418

Query Match	5.3%	Score 32.2;	DB 9;	Length 1185;
Best Local Similarity	49.1%;	Pred. No. 3.6;		
Matches	85;	Conservative	0;	Mismatches 88;
				Indels 0;
				Gaps 0;

Qy	382	ACCCTGCACTGCTGCTGCTTACAGTACAGTAAACGACAGATTATGCTGCCGACCCC	441
Db	265	ACCCTGACCTTGCCGCTGCTGCTGATACCTTCCTGACACTCTTGTTGGAGCAAGCG	324
Qy	442	TACACGACACACTGCTGCAGGCCCCCACTACGCGCTGTGTGCATGAATGCTTTGCG	501
Db	325	AATCTTCTGCTTTCTTCTGAGACAGCTCCAGCATGATGCGCCCTGGCGCGAAGCTTTGCGG	384
Qy	502	CCCTTGACCGATGCCAAGACTAGAGAGCATGCTGATGATGTGGGCTCTGCTCT	554
Db	385	GCGGTCACAGATGTGATCATTGCGAGCCAACTGGTGGAGAAAGTGTGCTTCT	437

```

1 RESULT 10
2 US-09-738-626-1/c
3 Sequence 1, Application US/09738626
4 Publication No. US20020197605A1
5
6 GENERAL INFORMATION:
7 APPLICANT: NAKAGAWA, SATOSHI
8 APPLICANT: MIZOGUCHI, HIROSHI
9 APPLICANT: ANDO, SEIKO
10 APPLICANT: HAYASHI, MIKIRO
11 APPLICANT: OCHIAI, KEIKO
12 APPLICANT: YOKOI, HARUHIKO
13 APPLICANT: TATEISHI, NAOKO
14 APPLICANT: SENOH, AKIHIRO
15 APPLICANT: IKEDA, MASATO
16 APPLICANT: OZAI, AKIO
17
18 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
19
20 FILE REFERENCE: 249-125
21
22 CURRENT APPLICATION NUMBER: US/09/738, 626
23
24 CURRENT FILING DATE: 2000-12-18
25
26 PRIOR APPLICATION NUMBER: JP 99/377484
27
28 PRIOR FILING DATE: 1999-12-16
29
30 PRIOR APPLICATION NUMBER: JP 00/159162
31
32 PRIOR FILING DATE: 2000-04-07
33
34 PRIOR APPLICATION NUMBER: JP 00/280988
35
36 PRIOR FILING DATE: 2000-08-03
37
38 NUMBER OF SEQ ID NOS: 7059
39
40 SOFTWARE: PatentIn ver. 3.0
41
42 SEQ ID NO 1
43
44 LENGTH: 3309400
45
46 TYPE: DNA
47
48 ORGANISM: Corynebacterium glutamicum
49
50 US-09-738-626-1

```

Query Match	5.3%	Score 32.2;	DB 9;	Length 3309400;
Best Local Similarity	49.1%;	Pred. No. 1.1e+02;		
Matches	85;	Conservative	0;	Mismatches 88;
			Indels	0;
			Gaps	0;

QY 382 ACCCTGACAGCTGCTGCTCCCTACAGTGAACAGTTAACTGCGGACCCC 441
Db 3301039 ACCGCTAACCTGCTCCGCTCTGCTGATGACCTCTTGCACAGTCTTGTGGTGAAGAACG 3300980
QY 442 TACCAACACAGACTTGCTCCAGCCCCCACTACGCGCTTGCTGCATGATGCTTTTTCG 501
Db 3300979 AATTTCTTCGTTCTTCTTGAGACACGCTCCACAGATGATGGCCCTGGCGGAACTCTTCGG 3300920
QY 502 CCCTTGACCGATGCCCAAGACTAGAGCCATGCTGATGATGTGGGTCCTGCTCT 554
Db 3300919 GCGGTGACCAATGCTGACACTGCGAGGAACTTGGTATGAGAAAGTTGCTTCCT 3300867

```

US-09-897-214-9
RESULT 11
; Sequence 9, Application US/09897214
; Patent No. US20020076779A1
; GENERAL INFORMATION:
; APPLICANT: Thayer, Edward C.
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Presnell, Scott R.
; TITLE OF INVENTION: Leucine-rich Repeat Proteins, Zlrr7,
; TITLE OF INVENTION: Zlrr8, and Zlrr9
; FILE REFERENCE: 01-27
; CURRENT APPLICATION NUMBER: US/09/897,214
; CURRENT FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 60/215,446
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1653
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate polynucleotide sequence
; NAME/KEY: misc feature
; LOCATION: (1) _(1653)
; OTHER INFORMATION: n = A,T,C or G
US-09-897-214-9

```

Query Match 5.2%; Score 31.8; DB 10; Length 1653;
Best Local Similarity 25.8%; Pred. No. 5.6;
Matches 60; Conservative 40; Mismatches 133; Indels 0; Gaps 0;

```
Oy 375 CCAGCCACCCCCTGCGCCTGTCTGCTTCAATGCACATTACGACGAGTTATGCTGC 434
      |||::||::||::||::||::||::||::||::||::||::||::||::||:
Db 390 YCARYYTNNGANNMNAATHGCCNCGNGCNTTGGAYGAYTTYTNGRBMNYNTGARGAYT 449
Oy 435 CGACCCCTCACCCACAACACTTGCTCCAGCCCCCACTAAGCGCGTGTGGCATGAATGC 494
      |||::||::||::||::||::||::||::||::||::||::||::||::||:
Db 450 NGAYTTMSNTAAIYAAYAAYTTMGNICARGTCNCITGGCGNGNATHSGNCGNAHTCCNGC 509
Oy 495 TTTTGCGGCTTGACCGATGCCAAGACTAGAGGCCATGCTGATGTGGGTCTTCGTTCT 554
      |||::||::||::||::||::||::||::||::||::||::||::||::||:
Db 510 MYTNCAVACNVTNAAAYYNGAYCAAAYTAATNTHAGVCANTNCCNCGNAGCNTTYGC 569
Oy 555 TTCTTCATTGAGGCTAGTATAATACCAAGGGGGAATPACACGGTTTTGCTCAT 607
      |||::||::||::||::||::||::||::||::||::||::||::||::||:
Db 570 NCARTYNGANCARITYNMNSMGNYTNGAIYYTACNMNSNAIYMNVTYNTGNCNACY 622
```

RESULT 12
US-09-815-242-7757/c
Sequence 7757, Application US/09815242/c
Patent No. US2002061569A1
GENERAL INFORMATION:
APPLICANT: Haeelbeck, Robert
APPLICANT: Onlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.

```

Query Match          5.3%; Score 32.2; DB 9; Length 1071;
Best Local Similarity 52.6%; Pred. No. 3.4;
Matches 70; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY      346 CAGGATGGTTACGCTGCATACCGCTACGCCCAACCCTTGCACACTGTGTCCTAC 405
DB      102 CATGCGCGTCCACACGGCGTGCGCCAGCCGGGCTCTCCGAGCCCCGTCGCCGCC 161
QY      406 AGTGACAAGTTACGGACGAGTTATGTGTGCCGACCCCTTACCAACACACTGTTCCAGCC 465
DB      162 TGCCACACTGACGGAACCGTCAACGTTGCCGCCCAAGCGCGCACACTCATCCTCCGG 221
QY      466 CCCACCTACGGCG 478
DB      222 CCCCgGTGTGACG 234

RESULT 8
US-09-880-505-180
Sequence 180, Application US/09880505
Publication No. US20030007976A1
GENERAL INFORMATION:
APPLICANT: Watson, James D.
APPLICANT: Tan, Paul L.J.
TITLE OF INVENTION: Methods and Compounds for the Treatment
FILE REFERENCE: 11000.1007c2
CURRENT APPLICATION NUMBER: US/09/880,505
PRIORITY FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 09/324,542
PRIOR FILING DATE: 1999-06-02
PRIOR APPLICATION NUMBER: US 08/997,080
NUMBER OF SEQ ID NOS: 194
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 180
LENGTH: 1071
TYPE: DNA
ORGANISM: Mycobacterium vaccae
US-09-880-505-180

Query Match          5.3%; Score 32.2; DB 9; Length 1071;
Best Local Similarity 52.6%; Pred. No. 3.4;
Matches 70; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY      346 CAGGATGGTTACGCTGCATACCGCTACGCCCAACCCTTGCACACTGTGTCCTAC 405
DB      102 CATGCGCGTCCACACGGCGTGCGCCAGCCGGGCTCTCCGAGCCCCGTCGCCGCC 161
QY      406 AGTGACAAGTTACGGACGAGTTATGTGTGCCGACCCCTTACCAACACACTGTTCCAGCC 465
DB      162 TGCCACACTGACGGAACCGTCAACGTTGCCGCCCAAGCGCGCACACTCATCCTCCGG 221
QY      466 CCCACCTACGGCG 478
DB      222 CCCCgGTGTGACG 234

RESULT 9
US-09-738-626-3418
Sequence 3418, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SEMOH, AKIHITO
APPLICANT: IKEDA, MASATO

```



```
Oy 301 GCGGAGTAGTGTATCAAGAGCCAGGTATGCGAATTAATGCTACAGGGTGTACGCT 360
    |||||
Db 835 GCGGAGTAGTGTATCAAGAGCCAGGTATGCGAATTAATGCTACAGGGTGTACGCT 894
Oy 361 GCATACCGCTACGCCAGCCAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
    |||||
Db 895 GCATACCGCTACGCCAGCCAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 954
Oy 421 CGAGTTTATGCTGCGCAGCCCTTACCAACACACTTGTCTCCAGCCCACTTACCGGCTT 480
    |||||
Db 955 CGAGTTTATGCTGCGCAGCCCTTACCAACACACTTGTCTCCAGCCCACTTACCGGCTT 1014
Oy 481 GGTGCATGATGCTTTTGCGCCCTTGACCGATGCGCAAGTACAGAGCCATG-CTGATGAT 540
    |||||
Db 1015 GGTGCATGATGCTTTTGCGCCCTTGACCGATGCGCAAGTACAGAGCCATG-CTGATGAT 1074
Oy 541 GTGGCTCTGCTTTCTTCTTCAATTCAGAGGCTAGTATATACCAAGGGGATACAAACGTTT 600
    |||||
Db 1075 GTGGCTCTGCTTTCTTCTTCAATTCAGAGGCTAGTATATACCAAGGGGATACAAACGTTT 1134
Oy 601 GCTCCATAT 609
    |||||
Db 1135 GCTCCATAT 1143
```

RESULT 2
US-09-794-591-1

```
; Sequence 1, Application US/09794591
; Patent No. US20010018198A1
; GENERAL INFORMATION:
; APPLICANT: Pulse, Stefan M.
; APPLICANT: Shibata, Hiroki
; TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
; TITLE OF INVENTION: Products Related Thereto and Methods of Using Same
; FILE REFERENCE: CE 3093
; CURRENT APPLICATION NUMBER: US/09/794,591
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/145,391
; PRIOR FILING DATE: 1998-09-01
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2372
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (987) .. (1979)
US-09-794-591-1
```

Query Match 70.4%; Score 428.6; DB 10; Length 2372;
Best Local Similarity 82.0%; Pred. No. 2.1e-125;
Matches 542; Conservative 0; Mismatches 64; Indels 55; Gaps 2;

```
Oy 1 ATGACTAATAAAGGCGGTGAACCCCTTACACCAATGGCTGAAATTAATTCAGTTGTG 60
    |||||
Db 1572 ATGACAAATAAAGGCGGTGAACCCCTTATACAAATGGCTGAAATTAATTCAGTTGTG 1631
Oy 61 GCGCGGCTACAGCCCGGCACTTCTATGACAGGACGGTGTCTGTGTGCCAGGCAACAG 120
    |||||
Db 1632 GGTGCAAGTACAGTCCCGAATTTCTATGACAGGACGGTGTCTGTGTGCCAGGCAACAG 1691
Oy 121 GAGGATCTTCCATGTAACAGTGGCCCACTTCACTTGTATATATCTTGTGAATGCTGCTG 180
    |||||
Db 1692 GAGGATCTTCCATGTAACAGTGGCCCACTTCACTTGTATATATCTTGTGAATGCTGCTG 1751
Oy 181 TTTCATATCCGCGCGCACTGCTGAGCTGACATACGAGGGGCTCACTTCAGAGCGCT 240
    |||||
Db 1752 TTTCATATCCGCGCGCACTGCTGAGCTGACATACGAGGGGCTCACTTCAGAGCGCGC 1811
Oy 241 GGTGCAACCGTGTACAAACACTTTCAGAGCTGCGCGCGCCACCCCAATCCGCGCTAT 300
    |||||
```

```
Db 1812 GGTGCAACCGTGTACAAACACTTTCAGAGCTGCGCGCGCCACCCCGATCCGCGCTAC 1871
Oy 301 GCGGAGTAGTGTATCAAGAGCCAGGTATGCGAATTAATGCTACAGGGTGTACGCT 360
    |||||
Db 1872 GCGGAGTAGTGTATCAAGAGCCAGGTATGCGAATTAATGCTACAGGGTGTACGCT 1929
Oy 361 GCATACCGCTACGCCAGCCAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 410
    |||||
Db 1930 GCATACCGCTACGCCAGCCAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1989
Oy 411 -----CAGTTACGAGAGGTT 427
    |||||
Db 1990 TTGCTCTTGTGACAGATGAAATTTCTTGAACACTCTGCAAGTACGAGAGGTT 2049
Oy 428 ATGTCGCGAGCCCTTACACCAACACACTTGTCTGACGCCCACTTACGCGGCTGTGCGCA 487
    |||||
Db 2050 ATGTCGCGAGCCCTTACACCAACACACTTGTCTGACGCCCACTTACGCGGCTGTGCGCA 2109
Oy 488 TGAATGCTTTTGCGCCCTTGAACCGATGCGCAAGACTAGAGCCATGCTATGATGTGCGTC 547
    |||||
Db 2110 TGAATGCTTTTGCGCCCTTGAACCGATGCGCAAGACTAGAGCCATGCTATGATGTGCGTC 2169
Oy 548 TCGTCTCTTCTTCAATTCAGAGGCTAGTATATACCAAGGGGATACAAACGTTTGTCTCAT 607
    |||||
Db 2170 TCGTCTCTTCTTCAATTCAGAGGCTAGTATATACCAAGGGGATACAAACGTTTGTCTCAT 2229
Oy 608 A 608
    |||||
Db 2230 A 2230
```

RESULT 3
US-10-184-644-202/C

```
; Sequence 202, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 202
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-644-202
```

Query Match 5.8%; Score 35.4; DB 9; Length 440;
Best Local Similarity 15.9%; Pred. No. 0.22;
Matches 52; Conservative 88; Mismatches 188; Indels 0; Gaps 0;

```
Oy 143 GCGCCAGTTCACTTGTATATCTTCTGCAATGCTTGGCTTTCCATATCCGCGCGCACAG 202
    |||||
Db 437 SSY.HMN.THMN..C.MSN.NMW.AH.N.KC.SA.SNBAHBCBTHCCSCWSSC.CYRN 378
Oy 203 CTGACAGTGCATACCGAGGGGCTCACTTTCAGAGCGCGTGTGCGACCGGTACCAACACT 262
    |||||
Db 377 HCSGC..YMC.MS..YKMSBC.CY.C..CS.CSEYT.NC.M.GC.MDNCSCSCDNCS 318
Oy 263 TCAGAGCTGCGCGCGCCACCCCAATCCGCGCTATGCGGAGTAGTATCAAGAGC 322
    |||||
```

GenCore version 5.1.4.g5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 12:44:58 ; Search time 54.057 Seconds
(without alignments)
7907.139 Million cell updates/sec

Title: US-09-809-545A-1_COPY_535_1143

Perfect score: 609
Sequence: 1 atgactataataaaagccgt.....acaacgcttcctccatat 609

Scoring table: IDENTITY NUC
Gapop 10'-0, Gapext 1.0

Searched: 501302 seqs, 350932545 residues

Total number of hits satisfying chosen parameters: 1002604

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubdna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubdna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubdna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubdna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubdna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubdna/PCTIS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubdna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubdna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubdna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubdna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubdna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubdna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubdna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubdna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	609	100.0	1340	10	US-09-809-545A-1
2	428.6	70.4	2372	10	US-09-794-591-1
3	35.4	5.8	440	9	US-10-184-644-202
4	34	5.6	735	9	US-10-184-644-88
5	32.2	5.3	653	9	US-10-184-644-402
6	32.2	5.3	951	9	US-09-738-626-3417
7	32.2	5.3	1071	9	US-10-051-643-180
8	32.2	5.3	1071	9	US-09-880-505-180
9	32.2	5.3	1185	9	US-09-738-626-3418
10	32.2	5.3	3309400	9	US-09-738-626-1
11	31.8	5.2	1653	10	US-09-897-214-9
12	31.2	5.1	927	10	US-09-815-242-7757
13	31.2	5.1	2558	9	US-09-557-232-3
14	31.2	5.1	9759	9	US-09-557-232-1
15	31	5.1	466	10	US-09-998-598-2571
16	31	5.1	1572	10	US-09-908-928-3
17	31	5.1	2127	10	US-09-908-928-1
18	31	5.1	15720	9	US-10-025-380-1058
19	31	5.1	15720	10	US-09-922-217-1058

C	20	31	5.1	15720	10	US-09-833-263-1058	Sequence 1058, Ap
C	21	30.6	5.0	671	9	US-10-184-644-346	Sequence 346, App
C	22	30.6	5.0	1459	10	US-09-964-824A-277	Sequence 277, App
C	23	30.4	5.0	678	9	US-10-184-644-150	Sequence 150, App
C	24	30.4	5.0	1141	9	US-10-184-644-120	Sequence 120, App
C	25	30.4	5.0	3642	10	US-09-969-515-7	Sequence 7, Appl
C	26	30.4	5.0	3675	10	US-09-969-515-1	Sequence 1, Appl
C	27	30.4	5.0	4666	10	US-09-969-515-13	Sequence 13, Appl
C	28	30.2	4.9	750	9	US-10-184-644-104	Sequence 104, App
C	29	30.2	4.9	403	10	US-09-960-352-8222	Sequence 8222, Ap
C	30	30	4.9	1873	9	US-09-986-480-94	Sequence 94, Appl
C	31	30	4.9	3639	9	US-10-012-896-779	Sequence 779, App
C	32	30	4.9	3639	9	US-09-895-793-779	Sequence 779, App
C	33	30	4.9	3639	9	US-09-895-814-779	Sequence 779, App
C	34	30	4.9	3639	10	US-09-759-143-779	Sequence 779, App
C	35	30	4.9	3639	10	US-09-780-669-779	Sequence 779, App
C	36	30	4.9	3639	10	US-09-822-827-779	Sequence 779, App
C	37	30	4.9	3918	10	US-09-781-311-1	Sequence 1, Appl
C	38	30	4.9	5668	9	US-10-012-896-777	Sequence 777, App
C	39	30	4.9	5668	9	US-09-895-793-777	Sequence 777, App
C	40	30	4.9	5668	9	US-09-895-814-777	Sequence 777, App
C	41	30	4.9	5668	10	US-09-759-143-777	Sequence 777, App
C	42	30	4.9	5668	10	US-09-780-669-777	Sequence 777, App
C	43	30	4.9	5668	10	US-09-822-827-777	Sequence 777, App
C	44	30	4.9	11869	10	US-09-764-869-2282	Sequence 2282, App
C	45	29.8	4.9	360	10	US-09-867-550-519	Sequence 519, App

ALIGNMENTS

RESULT 1
US-09-809-545A-1
Sequence 1, Application US/09809545A

Patent NO. US20020110804A1

GENERAL INFORMATION: Rattus norvegicus

APPLICANT: Stanton, Lawrence W.

APPLICANT: White, R. Tyler

TITLE OF INVENTION: SECRETED FACTORS

FILE REFERENCE: SCIOS.017A

CURRENT APPLICATION NUMBER: US/09/809,545A

CURRENT FILING DATE: 2001-03-14

NUMBER OF SEQ ID NOS: 84

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1

LENGTH: 1340

TYPE: DNA

ORGANISM: Rattus norvegicus

US-09-809-545A-1

Query Match 100.0%; Score 609; DB 10; Length 1340;
Best Local Similarity 100.0%; Pred. No. 1,4e-182;
Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGACTAATAAAGCCGCTGACACCCCTACACCAATGCTGAAATTAAATCCAGTTTG	60
DB	535	ATGACTAATAAAGCCGCTGACACCCCTACACCAATGCTGAAATTAAATCCAGTTTG	594
QY	61	GGCGGGCTACAGCCCGGACTTTCATGACGACGCGGCTGTGCGACGCAACAG	120
DB	595	GGCGGGCTACAGCCCGGACTTTCATGACGACGCGGCTGTGCGACGCAACAG	654
QY	121	GAGGATTTTCATGACGAGTGGCCCGGACTTTCATGATATATCTTGCAATGCTGGC	180
DB	655	GAGGATTTTCATGACGAGTGGCCCGGACTTTCATGATATATCTTGCAATGCTGGC	714
QY	181	TTTTCATATCCGCGCCGCTGCTGACGCTGATACGAGGGGCTCACTTCGAGCGCT	240
DB	715	TTTTCATATCCGCGCCGCTGCTGACGCTGATACGAGGGGCTCACTTCGAGCGCT	774
QY	241	GATGACCGGTGTACACACTTTCAGAGCTGGCGGCCCAACCCCAATCCGGCCAT	300
DB	775	GATGACCGGTGTACACACTTTCAGAGCTGGCGGCCCAACCCCAATCCGGCCAT	834

THIS PAGE BLANK (USPTO)

RESULT 14

US-08-997-080-180

Sequence 180, Application US/08997080

Patent No. 5985287

GENERAL INFORMATION:

APPLICANT: WATSON, JAMES D.

APPLICANT: TAN, PAUL E.J.

TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-

NUMBER OF SEQUENCES: 194

CORRESPONDENCE ADDRESS:

ADDRESSEE: Law Offices of Ann W. Speckman

STREET: 2601 Elliott Avenue, Suite 4185

CITY: Seattle

STATE: WA

COUNTRY: USA

ZIP: 98121

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/997,080

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Sleath, Janet

REGISTRATION NUMBER: 37,007

REFERENCE/DOCKET NUMBER: 11000.1007

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-269-0565

TELEFAX: 206-269-0563

TELEX:

INFORMATION FOR SEQ ID NO: 180:

SEQUENCE CHARACTERISTICS:

LENGTH: 1071 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-997-080-180

Query Match 5.3%; Score 32.2; DB 2; Length 1071;

Best Local Similarity 52.6%; Pred. No. 2.2;

Matches 70; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 346 CAGGATGTTACGCTGATACCGCTACGCCAGCCACCCCTGCACTGCTGCTTAC 405

DB 102 CATCGCGTCCACACGCGGTGCGCCAGCGGCTCTCCGCGAGCCCGCTGCGCGCC 161

QY 406 AGTACAGTTACGACGAGATTATGCTGCCGACCCCTACCAACACACTTGTCCAGCC 465

DB 162 TGCCACAGTACGCAACCGTACAGTTGCGCCCAAGCGCGGCAACAATCATCCCGCG 221

QY 466 CCCACCTACGCGC 478

DB 222 CCGCGTGTGACG 234

RESULT 15

US-08-997-362-180

Sequence 180, Application US/08997362

Patent No. 5985287

GENERAL INFORMATION:

APPLICANT: Tan, Paul

APPLICANT: Hiyama, Jun

APPLICANT: Visser, Elizabeth

APPLICANT: Skinner, Margot

APPLICANT: Scott, Linda

APPLICANT: Prestidge, Ross

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR

TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS

NUMBER OF SEQUENCES: 194

CORRESPONDENCE ADDRESS:

ADDRESSEE: Law Offices of Ann W. Speckman

STREET: 2601 Elliott Avenue, Suite 4185

CITY: Seattle

STATE: WA

COUNTRY: USA

ZIP: 98121

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/997,362

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970

FILING DATE: June 12, 1997

APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347

FILING DATE: August 29, 1996

ATTORNEY/AGENT INFORMATION:

NAME: Sleath, Janet

REGISTRATION NUMBER: 37,007

REFERENCE/DOCKET NUMBER: 11000.1002c2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-269-0565

TELEFAX: 206-269-0563

TELEX:

INFORMATION FOR SEQ ID NO: 180:

SEQUENCE CHARACTERISTICS:

LENGTH: 1071 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-997-362-180

Query Match 5.3%; Score 32.2; DB 2; Length 1071;

Best Local Similarity 52.6%; Pred. No. 2.2;

Matches 70; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 346 CAGGATGTTACGCTGATACCGCTACGCCAGCCACCCCTGCACTGCTGCTTAC 405

DB 102 CATCGCGTCCACACGCGGTGCGCCAGCGGCTCTCCGCGAGCCCGCTGCGCGCC 161

QY 406 AGTACAGTTACGACGAGATTATGCTGCCGACCCCTACCAACACACTTGTCCAGCC 465

DB 162 TGCCACAGTACGCAACCGTACAGTTGCGCCCAAGCGCGGCAACAATCATCCCGCG 221

QY 466 CCCACCTACGCGC 478

DB 222 CCGCGTGTGACG 234

Search completed: March 15, 2003, 16:22:18

Job time: 1277.68 secs

LOCATION: 15429.15664
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: /frame"
US-08-402-068-3

5.6%; Score 34.2; DB 1; Length 15664;

Best Local Similarity 47.8%; Pred. No. 1.8; Mismatches 108; Indels 0; Gaps 0;

Query March
Matches 99; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

Qy 220 GGGGCTACCTTCGAGCGCGGTGTCACCGTGTACAAACCTTCAGAGTGGCGGCC 279
Db 720 GGGGCTACCTTCGAGCGCGGTGTCACCGTGTACAAACCTTCAGAGTGGCGGCC 779
Qy 280 CCACCCCAATCCGGCTATGGCGGAGTGTATCAAGAGCAAGTATGCAATAA 339
Db 780 GCCGTGCTATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 839
Qy 340 TTGCTACAGGGGTGTTACGTCATACCGCTACCGCCGACCGCCGCGCGCGCGCT 399
Db 840 CTGACTTGTGCGGAGCGCGCGGTTCGGCTATGACAGCGCGCGCGGTATGTCGA 899
Qy 400 GCCTACAGTACAGTACGAGAGATT 426
Db 900 GCCACACCGTGTCCAGCCCGCAT 926

RESULT 10

US-09-056-180/c
Sequence 180, Application US/09056556
Patent No. 6350456

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, David C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND

NUMBER OF SEQUENCES: 241

CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle

STATE: Washington
COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556

FILING DATE: 07-APR-1998
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.

REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457

TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 180:

SEQUENCE CHARACTERISTICS:
LENGTH: 538 base pairs

TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear

US-09-056-180

Query Match 5.5%; Score 33.4; DB 4; Length 538;
Best Local Similarity 51.0%; Pred. No. 0.68; Mismatches 76; Indels 0; Gaps 0;

Matches 79; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 357 CGCTGATACCGGTACCGCCGAGCCCGCATGCTGCTGCTGCTGCTGCTGCTGCT 416
Db 143 CGCGCTGCGCGGTGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 109

Db 263 CGTACCGCGGTACCGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 204

Qy 417 CGGACGATTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 476

Db 203 CGCGCTGCGCGGTGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 144

Qy 477 CGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 511

Db 143 CGCGCTGCGCGGTGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 109

RESULT 11

US-09-072-596-175/c
Sequence 175, Application US/09072596
Patent No. 6458366

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonia

APPLICANT: Houghton, Raymond
APPLICANT: Vedick, Thomas S.

APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
NUMBER OF SEQUENCES: 350

CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle

STATE: Washington
COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596

FILING DATE: 05-MAY-1998
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.

REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9

TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 175:

SEQUENCE CHARACTERISTICS:
LENGTH: 538 base pairs

TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear

US-09-072-596-175

Query Match 5.5%; Score 33.4; DB 4; Length 538;
Best Local Similarity 51.0%; Pred. No. 0.68; Mismatches 76; Indels 0; Gaps 0;

Matches 79; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 357 CGCTGATACCGGTACCGCCGAGCCCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 416

Db 263 CGTACCGCGGTACCGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 204

Qy 417 CGGACGATTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 476

Db 203 CGCGCTGCGCGGTGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 144

Qy 477 CGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 511

Db 143 CGCGCTGCGCGGTGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 109

CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
ADDRESS: Company
STREET: 1 Becton Drive
CITY: Franklin Lakes
STATE: NJ
COUNTRY: US
ZIP: 07417
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/402,068
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fugit, Donna R.
REGISTRATION NUMBER: 32,135
REFERENCE/DOCKET NUMBER: P-3283
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 15664 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 222..425
OTHER INFORMATION: /function= "potential open reading."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 451..747
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 747..1109
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1109..2014
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2034..2747
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2747..3109
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3109..3444
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3444..3728
OTHER INFORMATION: /function= "potential open reading"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3731..4855
OTHER INFORMATION: /function= "potential open reading"

FEATURE:
NAME/KEY: misc_feature
LOCATION: 4855..5376
OTHER INFORMATION: /function= "potential coding"
OTHER INFORMATION: sequence"
OTHER INFORMATION: /product= "L5 gp37 homolog"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5382..5747
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5837..6307
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6403..7770
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7770..8006
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 8033..8236
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 8244..9443
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9450..10244
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10371..10586
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1115..11786
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11917..12741
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12748..14499
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14771..15154
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15154..15426
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature

```
NAME/KEY: misc_feature
LOCATION: 747..1109 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1109..2014 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2034..2747 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2747..3109 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3109..3444 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3444..3728 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3731..4855 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4855..5376 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5382..5747 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5837..6307 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6403..7770 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7770..8006 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 8033..8236 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 8244..9443 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9450..10244
```

```
OTHER INFORMATION: /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10371..10586 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11115..11786 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11917..12741 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12748..14499 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14771..15154 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15154..15426 /function= "potential open reading
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15429..15664 /function= "potential open reading
OTHER INFORMATION: frame"
US-08-402-066-3
Query Match 5.6%; Score 34.2; DB 1; Length 15664;
Best Local Similarity 47.8%; Pred. No. 1.8;
Matches 99; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
QY 220 GGGGCTACCTTGAGGCGGTGTCACCGTGTGACACACCTTGAGAGCTGGCGGCC 279
DB 720 GCGGCTTACTGGAGGCGGTGTCGATGATCCGCGAGCAATGCAATCGGCGTGGCC 779
QY 280 CCACCCCAATCCGGCCTATGCGGAGTAGTGTATCAAGGCCAGTGTATGGCAATAA 339
DB 780 GCCGTGCTATGCGCGCGGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCG 839
QY 340 TTGCTACAGGGGTGTATGACCTGATACCGCCAGCCGCCAGCCCTGCGACTGCTCT 399
DB 840 CTGTACTTCTCGCGGAGCGCGGTTCCGCTATATGACAGCGCGCGCGCGCTGCTCCA 899
QY 400 GCGTACAGTGACAGTTACGAGCAAGTT 426
DB 900 GGCACACCGTGTGCCAAGCCCGGCGAT 926
RESULT 9
US-08-402-068-3
Sequence 3, Application US/08402068
Patent No. 5633159
GENERAL INFORMATION:
APPLICANT: Pearson, Robert E.
APPLICANT: Dickson, Julie A.
APPLICANT: Hamilton, Paul T.
APPLICANT: Little, Michael C.
APPLICANT: Beyer Jr., Wayne F.
TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX
NUMBER OF SEQUENCES: 6
```

OTHER INFORMATION: /product= "L5 gp37 homolog"
FEATURE:
NAME/KEY: misc feature
LOCATION: 5382..5747
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 5837..6307
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 6403..7770
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 7770..8006
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 8033..8236
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 8244..9443
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 9450..10244
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 10371..10586
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 11115..11786
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 11917..12741
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 12748..14499
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 14771..15154
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 1515f..15426
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 15425..15664
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
US-08-508-004-3

Query Match 5.6%; Score 34.2; DB 1: Length 15664;
Best Local Similarity 47.8%; Pred. No. 1.8;
Matches 99; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
QY 220 GGGGCTACCTTCGAGGCGTGTGTGCGACCGTGTACACACCTTCAGAGCTGCGGCC 279
DB 720 GGGGCTACTGAGAGGCGGTCTGCTGATGATCCGCGACGAATGCAATGGGCTGGCC 779
QY 280 CCACCCCAATCCGCGCTATGCGGAGTGTATCAAGACCAAGTATGGCAATAA 339
DB 780 GCCGTGCTATGCGGCGCGCGCGCGCGCGCGCAACCGCAACAGACGCGCTTC 839
QY 340 TTGCTACAGGTTGTATGCTGATACCGCTACCGCCAGCCCGCTCCACTGCTCT 399
DB 840 CTGTACTTGTCTGCGGAGCGCGGTTCGCTATGAGAGCGCGCGGTATGCTGCA 899
QY 400 GCCTACAGTGAAGTTACGACGAGTT 426
DB 900 GGCACACCGTGTGCCAAGCCCGCAT 926

RESULT 8
US-08-402-066-3
Sequence 3, Application US/08402066
Patent No. 5612182
GENERAL INFORMATION:
APPLICANT: Pearson, Robert E.
APPLICANT: Dickson, Julie A.
APPLICANT: Hamilton, Paul T.
APPLICANT: Little, Michael C.
APPLICANT: Beyer Jr., Wayne F.
TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
ADDRESSEE: Company
STREET: 1 Becton Drive
CITY: Franklin Lakes
STATE: NJ
COUNTRY: US
ZIP: 07417
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/402,066
FILING DATE:
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Pugt, Donna R.
REGISTRATION NUMBER: 32,135
REFERENCE/DOCKET NUMBER: P-3283
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 15664 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc feature
LOCATION: 222..425
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:
NAME/KEY: misc feature
LOCATION: 451..747
OTHER INFORMATION: /function= "potential open reading"
OTHER INFORMATION: frame"
FEATURE:

```
NAME/KEY: misc_feature
LOCATION: 11115..11786
OTHER INFORMATION: /function= "potential open reading
FRAME"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11917..12741
OTHER INFORMATION: /function= "potential open reading
FRAME"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12748..14499
OTHER INFORMATION: /function= "potential open reading
FRAME"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14771..15154
OTHER INFORMATION: /function= "potential open reading
FRAME"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15154..15426
OTHER INFORMATION: /function= "potential open reading
FRAME"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15429..15664
OTHER INFORMATION: /function= "potential open reading
FRAME"
OTHER INFORMATION: "frame"
```

US-08-402-282-3

```
Query Match 5.6%; Score 34.2; DB 1; Length 15664;
Best Local Similarity 47.8%; Pred. No. 1.8;
Matches 99; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
```

```
OY 220 GGGGTCACCTTCAGGCGCGTGTGCGACCGGTGACACCTTACAGGCTGCGCGCC 279
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 720 GCGCGTCTACTGCGAGGCGGCTGCTGCGATGATCCCGCAGCATGCGATCGGCTGCGC 779
OY 280 CCACCCCAATCCCGCGCTATGCGAGTAGTATCAAGACGAGTATGATGCAATAA 339
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 780 GCGGTGCTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 839
OY 340 TTGCTACAGGAGTGTACGCTGATACCGCTACGCGCGCGCGCGCGCGCGCGCGCT 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 840 CTGTAAGTGTCTGCGAGGCGCGGCTTGGCTATGACAGCGCGCGCGCGCTGCTCGA 899
OY 400 GCCTACAGTACAGTACGAGCGAGTT 426
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 900 GGCACACCGGTGTGCCAAGCCCGCGAT 926
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 7

```
US-08-508-004-3
Sequence 3, Application US/08508004
Patent No. 5582969
```

GENERAL INFORMATION:

```
APPLICANT: Pearson, Robert E.
APPLICANT: Dickson, Julie A.
APPLICANT: Hamilton, Paul T.
APPLICANT: Little, Michael C.
APPLICANT: Beyer Jr., Wayne F.
TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: Richard J. Rodrick, Becton Dickinson and
ADDRESS: Company
STREET: 1 Becton Drive
CITY: Franklin Lakes
STATE: NJ
COUNTRY: US
ZIP: 07417
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,004
FILING DATE: 27-JUL-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/402,282
FILING DATE: 10-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fugit, Donna R.
REGISTRATION NUMBER: 32,135
REFERENCE/DOCKET NUMBER: P-3283
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 15664 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 222..425
OTHER INFORMATION: /function= "potential open reading
FRAME"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 451..747
OTHER INFORMATION: /function= "potential open reading
FRAME"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 747..1109
OTHER INFORMATION: /function= "potential open reading
FRAME"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1109..2014
OTHER INFORMATION: /function= "potential open reading
FRAME"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2034..2747
OTHER INFORMATION: /function= "potential open reading
FRAME"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2747..3109
OTHER INFORMATION: /function= "potential open reading
FRAME"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3109..3444
OTHER INFORMATION: /function= "potential open reading
FRAME"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3444..3728
OTHER INFORMATION: /function= "potential open reading
FRAME"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3731..4855
OTHER INFORMATION: /function= "potential open reading
FRAME"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4855..5376
OTHER INFORMATION: /function= "potential coding
sequence"
```

```

Oy 404 ACAGGACAGTTACGACAGAGTTTATGCTGGCCGACCCCTACACACACACACTTGTCCGAG 463
Db 3942479 ATGCCAGACACCCCGCGCCCGCTTGTCTCCGCGGCCCGCGGACCGGACCTTGTGTG 39424220
Oy 464 CCCCCACCTTACGGCGTGTGTCATGAATGCTTTTGCGCCCTTTGACCGATGCGCAAGACTA 523
Db 3942419 CCGCGCGGTGCGCGGCTTCCCGCGGCCCGCGCTTGGCGCGGGGCGCGGCAAGCGG 39423600
Oy 524 GG 525
Db 3942359 GG 3942358

RESULT 6
US-08-402-282-3
: Sequence 3, Application US/08402282
: Patent No. 5476768
: GENERAL INFORMATION:
: APPLICANT: Pearson, Robert E.
: APPLICANT: Dickson, Julie A.
: APPLICANT: Hamilton, Paul T.
: APPLICANT: Little, Michael C.
: APPLICANT: Beyer Jr., Wayne F.
: TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
: TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
: ADDRESSEE: Company
: STREET: 1 Becton Drive
: CITY: Franklin Lakes
: STATE: NJ
: COUNTRY: US
: ZIP: 07417
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: . OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/402,282
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Fugit, Donna R.
: REGISTRATION NUMBER: 32,135
: REFERENCE/DOCKET NUMBER: P-3283
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 15664 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 451..747
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 747..1109
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 1109..2014
: OTHER INFORMATION: /function= "potential open reading

```

OTHER INFORMATION: frame"	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: 2034..2747	
OTHER INFORMATION: /function=	"potential open reading
OTHER INFORMATION: frame"	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: 2747..3109	
OTHER INFORMATION: /function=	"potential open reading
OTHER INFORMATION: frame"	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: 3109..3444	
OTHER INFORMATION: /function=	"potential open reading
OTHER INFORMATION: frame"	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: 3444..3728	
OTHER INFORMATION: /function=	"potential open reading
OTHER INFORMATION: frame"	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: 3731..4855	
OTHER INFORMATION: /function=	"potential open reading
OTHER INFORMATION: frame"	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: 4855..5376	
OTHER INFORMATION: /function=	"potential coding
OTHER INFORMATION: sequence"	
OTHER INFORMATION: /product=	"L5 gp37 homolog"
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: 5382..5747	
OTHER INFORMATION: /function=	"potential open reading
OTHER INFORMATION: frame"	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: 5837..6307	
OTHER INFORMATION: /function=	"potential open reading
OTHER INFORMATION: frame"	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: 6403..7770	
OTHER INFORMATION: /function=	"potential open reading
OTHER INFORMATION: frame"	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: 7770..8006	
OTHER INFORMATION: /function=	"potential open reading
OTHER INFORMATION: frame"	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: 8033..8236	
OTHER INFORMATION: /function=	"potential open reading
OTHER INFORMATION: frame"	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: 8244..9443	
OTHER INFORMATION: /function=	"potential open reading
OTHER INFORMATION: frame"	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: 9450..10244	
OTHER INFORMATION: /function=	"potential open reading
OTHER INFORMATION: frame"	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: 10371..10586	
OTHER INFORMATION: /function=	"potential open reading
OTHER INFORMATION: frame"	
FEATURE:	

```
Db 1078 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1137
Qy 187 TATCGGCGGCGCTGCTGAGCATATACGAGGGGCGTCACTGAGCGGTGTCGC 246
Db 1138 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1197
Qy 247 ACCGTGACACACCTTGAGAGCTGGCGGCCCGCCCAATCCGCGCTTATGCGGA 306
Db 1198 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1257
Qy 307 GTAGTATCAAGACCGAGTATGAGCATTAATGCTACAGAGGTGTTACGCTGATAC 366
Db 1258 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1317
Qy 367 CGTAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 426
Db 1318 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1377
Qy 427 TATGCTGCGGACCCCTACACACACACTGCTCCAGCCCGCCACTGAGCGGTGTC 486
Db 1378 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1437
Qy 487 ATGAATGCTTTGGCGCCCTTGACCGATGCCAGACTAGG 525
Db 1438 CCAATCTCTATCTCTTAATACTGACATAGATAGG 1476
```

```
RESULT 4
US-08-323-443B-1
Sequence 1, Application US/08323443B
Patent No. 5654170
GENERAL INFORMATION:
APPLICANT: KLINGER, KATHERINE W.
APPLICANT: LANDES, GREGORY M.
APPLICANT: BURN, TIMOTHY C.
APPLICANT: CONNORS, TIMOTHY D.
APPLICANT: DACKOWSKI, WILLIAM R.
APPLICANT: GERMINO, GREGORY
APPLICANT: QIAN, FENG
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323,443B
FILING DATE: 12-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, S. Peter
REGISTRATION NUMBER: 25,351
REFERENCE/DOCKET NUMBER: 0372/0A462
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 527-7700
TELEFAX: (212) 753-6237
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 31571 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
```

```
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: PKD1 GENOMIC
US-08-323-443B-1
Query Match 6.0%; Score 36.8; DB 1; Length 31571;
Best Local Similarity 47.3%; Pred. No. 0.39;
Matches 95; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
Qy 52 CCAGTGTGGGCGGCGGCTACAGCCCGGACTTATGAGGACAGGAGGCTGTGGCCAG 111
Db 25427 CCAGTGTGTGTGAGACATGTATGACACACTAGCGTCCGNNACTGGGTNAGNAGG 25486
Qy 112 GCCAACGAGGAGGATCTTCATGATAGTGGCCCGCACTTCACTGTATATATCTTGCA 171
Db 25487 GCNAGNNNTGGGAGNAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 25546
Qy 172 ATGCTGCTTTCATATCCGCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 231
Db 25547 GGGCTGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 25606
Qy 232 CGAGCGGTGTGTGCGACCGTG 252
Db 25607 CTCGCGAGGACGCCCTCAGTG 25627
```

```
RESULT 5
US-09-103-840A-2/C
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
```

```
Query Match 5.8%; Score 35.6; DB 4; Length 4403765;
Best Local Similarity 43.6%; Pred. No. 8;
Matches 158; Conservative 0; Mismatches 204; Indels 0; Gaps 0;
Qy 164 CTTGCAATGCGCTGGCTTTCATATCCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 223
Db 3942719 CTGCGCGGCGGCGGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3942660
Qy 224 CTCACCTTCGAGCGGCTGTGCGACCGGTATCAACACCTTCAGAGCTGCGGCGGCCAC 283
Db 3942659 CCGTTGAAGTTGCGCCCGCGGCTACCGCGGCGGCGGCGGCGGCGGCGGCGG 3942600
Qy 284 CCCCAATCCGCGCTTATGCGGAGTATGATATCAAGAGCCAGTATATGCAATTAATTC 343
Db 3942599 CCGCGGCGGCGGCTTTCGCGGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGG 3942540
Qy 344 TACAGGATGTAGCTGTGCAATACCGCTACCGCCAGCCACCCCGTGCACGCTGCTGCT 403
Db 3942539 CCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3942480
```


Qy	361	GCATCCGCTAACGCCACGAGCCACCCCGGACGCTGCTGCTCAAGTGA-----	410
Db	1930	GCATACCGCTAACGCCACGCTACCCCTGCACCTGGCTTACAGTGACGAATATAG	1989
Qy	411	-----CAGTTACGGACAGATT	427
Db	1990	TTTCGCTTTTCGTTGACGACGATGAAATTTCTTGTACACCTCTGCAGTTACGGACAGATT	2049
Qy	428	ATGCTGCGGACCCCTTACACCAACAACATTGCTTCAGCCCCCACTTACGGCGTTTGCTGCA	487
Db	2050	ATGCTGCGGACCCCTTACACCAACGCACTTGCTTCAGCCCCCACTTACGGCGTTTGCTGCA	2109
Qy	488	TGAATGCTTTTGGCGCCTTTCAGCCGATGCCAAGACTAGAGACATGCTGATGATGGGCT	547
Db	2110	TGAATGCTTTTGGCACCCTTTCAGTCCGATGCCAAGACTAGAGACATGCTGATGATGGGCT	2169
Qy	548	TGCTTTCTTTCTTCAATTGACAGGCTATGATATACCAAGGGGATATCAACGTTTTCCTCAT	607
Db	2170	TGCTTTCTTTCTTCAATTGACAGGCTATGATATACCAAGGGGATATCAACCGTTTTCCTCAT	2239
Qy	608	A 608	
Db	2230	A 2230	

```

RESULT 2
US-09-176-657-5
Sequence 5, Application US/09176657
Patent No. 6020164
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Tang, Y. Tom
APPLICANT: Corley, Neil C.
APPLICANT: Guegler, Karl J.
APPLICANT: Lu, Aina
APPLICANT: Baughn, Mariah R.
TITLE OF INVENTION: HUMAN RNA BINDING PROTEINS
FILE REFERENCE: PF-0611 US
CURRENT APPLICATION NUMBER: US/09/176,657
CURRENT FILING DATE: 1998-10-21
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PERL Program
SEQ ID NO 5
LENGTH: 1506
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE: -
OTHER INFORMATION: 1250374
US-09-176-657-5

Query Match          10.4%; Score 63.2; DB 3; Length 1506;
Best Local Similarity 54.5%; Pred. No. 7, 2e-10;
Matches 279; Conservative 0; Mismatches 193; Indels 40; Gaps

QY      1 ATGACTAATATAAAAGGCCGTGAACCCCTTACCAATGGCTGGAAATTAAATTCAGTTTG 60
        ||| |||| |
DB       768 ATGACCAAATTAAGAAGTGTGCACCATATGCAAAATGGTTGAATTAAGCCAGTAGTT 827
        ||| |||| |
QY      61 GGCGCGGTATACAGCCCCGCACTTCTATGAGCAGCACGGTGTCT-----TGTCCAG 111
        ||| |||| |
DB       828 GGAAGCTGTATATGTGTCCGAGGTATATGACAGATCCAGCTTTCAAGACAGATGTGTCCCTA 887
        ||| |||| |
QY      112 GCCAACCCAGAGGAGATCTTCATGTACATGAGTGGCCCCCACTTCACTTGTATTACTTCTGCA 171
        ||| |||| |
DB       888 GGCMAATGATGACACACAGTGCCTCATAGGAAGAGGGGGATATCAACCTTAACATTTCTTTA 947
        ||| |||| |
QY      172 ---ATGCTTGGCTTTTGCATA---TCCGGCGCGCCACTGCTSCAGCTCATACGAGGGGCT 225
        ||| |||| |
DB       948 ATCATTTCTGTGCTTCCCTTAACCTTACGTGCAGCACAACGAGCAGCCCTTTCAAGAGAGCC 1007
        ||| |||| |
QY      226 CACCTTGCAGGCGGTGTGCGACCGTGTACACAACCTTCAGAGCTGGGCGGCCCTCACCC 285
        ||| |||| |

```

Db	1008	CAATTGAGGGGACAGAGGGGCGGACAGATATATGGTCAGTCCGAAA---GGGTAACCTCCAAACA	1064
Qy	286	CCAAATCCCGGCTATATGCGCGAGTAGTAGTATCAAGAGCCAGTGTATGGCAATAAATTGCTA	345
Db	1065	GGCAATCCCCCGCTATATCCAGGGGGTGGATATAGCAGCCCTACAGATATGCACAGCCTGCTACTG	1124
Qy	346	CAGAGTGGTATACGCTGCATACCGCTACGGCCAGCCACCCTCGCACTGCTCTCTCTAC	405
Db	1125	CACAC-----CGAGGCACCGCTGCTGTGCAGCCGCTGTGCAGCCGCTTAC	1165
Qy	406	AGTAGACATTACGACAGAGTTTATGTCTGCCAGCCCTTACACACACAACTTGTCTCCACC	465
Db	1166	AGTAGCGTTATGCGACAGGGTGTACACAGCCGCAACCCCT---ACCAATGCCCTTGCCTCCCTCC	1222
Qy	466	CCCACTACAGCGCGTTGGTGTGCATGAAATGCTTT	497
Db	1223	GCTACCTATGAGATTGGCGCTGTGGCGAAGTTT	1254

```

127 TCTTCATGTAAGAGTGCCCGCATTCATTGATATACCTTGCGAATGCGCTGCTTCCA 186
Query Match 6.1%; Score 37.4; DB 1; Length 7218;
Best Local Similarity 5.0%; Pred. No. 0.13;
Matches 20; Conservative 204; Mismatches 175; Indels 0; Gaps 0

```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 12:42:03 ; Search time 34.6839 Seconds
(without alignments)
5384.802 Million cell updates/sec

Title: US-09-809-545A-1_COPY_535_1143
Perfect score: 609
Sequence: 1 atgacataataaaggccgtt.....acaacgtcttcctccatcat 609

Scoring table: IDENTITY NUC
Gapop 10'-0', Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PTCUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	428.6	70.4	2372	4 US-09-145-391-1	Sequence 1, Appli
2	63.2	10.4	1506	3 US-09-176-657-5	Sequence 5, Appli
3	37.4	6.1	7218	1 US-08-232-463-14	Sequence 1, Appli
4	36.8	6.0	31571	1 US-08-323-443B-1	Sequence 2, Appli
5	35.6	5.8	4403765	4 US-09-103-840A-2	Sequence 3, Appli
6	34.2	5.6	15664	1 US-08-402-282-3	Sequence 3, Appli
7	34.2	5.6	15664	1 US-08-508-004-3	Sequence 3, Appli
8	34.2	5.6	15664	1 US-08-402-066-3	Sequence 3, Appli
9	34.2	5.6	15664	1 US-08-402-068-3	Sequence 3, Appli
10	33.4	5.5	538	4 US-09-056-556-180	Sequence 180, App
11	33.4	5.5	538	4 US-09-072-596-175	Sequence 175, App
12	32.8	5.4	535	4 US-09-056-556-171	Sequence 171, App
13	32.8	5.4	535	4 US-09-072-596-166	Sequence 166, App
14	32.2	5.3	1071	2 US-08-997-080-180	Sequence 180, App
15	32.2	5.3	1071	2 US-08-997-362-180	Sequence 180, App
16	32.2	5.3	1071	4 US-09-095-855-180	Sequence 180, App
17	32.2	5.3	1071	4 US-09-324-542-180	Sequence 180, App
18	32.2	5.3	1071	4 US-09-205-426-180	Sequence 180, App
19	31.8	5.2	610	4 US-09-072-596-293	Sequence 293, App
20	31.6	5.2	418	3 US-09-141-000-6	Sequence 6, Appli
21	31.4	5.2	458	3 US-09-141-000-4	Sequence 4, Appli
22	31.4	5.2	441529	4 US-09-103-840A-1	Sequence 1, Appli
23	31.2	5.1	2558	3 US-08-999-733-3	Sequence 3, Appli
24	31.2	5.1	9759	1 US-08-093-453B-1	Sequence 1, Appli
25	31.2	5.1	9759	1 US-08-453-041A-1	Sequence 1, Appli
26	31.2	5.1	9759	3 US-08-999-733-1	Sequence 1, Appli
27	30.8	5.1	441529	4 US-09-103-840A-1	Sequence 1, Appli

C	28	30.6	5.0	4793	4 US-09-561-497-10	Sequence 10, Appli
	29	30	4.9	3639	4 US-09-605-785-779	Sequence 779, App
	30	30	4.9	3848	4 US-09-112-096-28	Sequence 28, Appli
	31	30	4.9	5668	4 US-09-112-096-14	Sequence 14, Appli
	32	30	4.9	5668	4 US-09-605-785-777	Sequence 777, App
	33	29.8	4.9	744	4 US-09-385-028-22	Sequence 22, Appli
	34	29.8	4.9	2283	4 US-09-153-804-5	Sequence 5, Appli
	35	29.8	4.9	2903	2 US-08-310-912A-1	Sequence 1, Appli
	36	29.8	4.9	2903	3 US-08-841-089-1	Sequence 1, Appli
	37	29.8	4.9	2903	4 US-09-301-085-1	Sequence 1, Appli
	38	29.8	4.9	2903	5 PCT-US95-04570-1	Sequence 1, Appli
	39	29.8	4.9	2903	5 PCT-US95-04589-1	Sequence 1, Appli
	40	29.8	4.9	3432	2 US-08-310-912A-141	Sequence 141, App
	41	29.8	4.9	3432	4 US-09-301-085-141	Sequence 141, App
	42	29.8	4.9	3432	5 PCT-US95-04589-141	Sequence 141, App
	43	29.8	4.9	11604	4 US-09-385-028-13	Sequence 13, Appli
	44	29.8	4.9	15079	4 US-09-385-028-1	Sequence 1, Appli
	45	29.6	4.9	11673	4 US-09-334-220-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1									
US-09-145-391-1									
; Sequence 1, Application US/09145391									
; Patent No. 6194171									
; GENERAL INFORMATION:									
; APPLICANT: Pulat, Stefan M.									
; APPLICANT: Shiba, Hiroki									
; TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,									
; FILE REFERENCE: CE 3093									
; CURRENT APPLICATION NUMBER: US/09/145.391									
; CURRENT FILING DATE: 1998-09-01									
; NUMBER OF SEQ ID NOS: 2									
; SOFTWARE: PatentIn Ver. 2.0									
; SEQ ID NO 1									
; LENGTH: 2372									
; TYPE: DNA									
; ORGANISM: Homo sapiens									
; FEATURE:									
; NAME/KEY: CDS									
; LOCATION: (987)..(1979)									
US-09-145-391-1									
Query Match									
Best Local Similarity 70.4%; Score 428.6; DB 4; Length 2372;									
Matches 542; Conservative 0; Mismatches 64; Indels 55; Gaps 2;									
QY	1	ATGACTAATAAAGGCGCGTGAACCCCTACACCAATGCGGAATTAATCCAGTTGTG	60						
DB	1572	ATGACAAATTAATAAGACCGCTACCCCTTATACAAATGCTGGAATTAATCCAGTTGTG	1631						
QY	61	GCGCGGTCTACAGCGCCGACTTCTATGACGACGCGTCTGTGTGCGAGGCCAACAG	120						
DB	1632	GGTGAGCTACAGCGCCGAAATTCATGACGACGCGTCTGTGTGCGAGGCCAACAG	1691						
QY	121	GAGGATTTTCATGATGATGAGGCGCCGACTTCTATGATTAATCTTCGCAATGCTGGC	180						
DB	1692	GAGGATTTTCATGATGATGAGGCGCCGACTTCTATGATTAATCTTCGCAATGCTGGC	1751						
QY	181	TTTCATATTCGCGCGCGCCGCTGCTGATGATGATGATGATGATGATGATGATGATG	240						
DB	1752	TTTCATATTCGCGCGCGCCGCTGCTGATGATGATGATGATGATGATGATGATGATG	1811						
QY	241	GATGCGACCGGTGATCAACACCTTCAGAGCTCGCGGCCGCCACCCCAATCCGGCTAT	300						
DB	1812	GATGCGACCGGTGATCAACACCTTCAGAGCTCGCGGCCGCCACCCCAATCCGGCTAT	1871						
QY	301	GCGGAGTGTGTATCAAGAGCGCGATGATGATGATGATGATGATGATGATGATGAT	360						
DB	1872	GCGGAGTGTGTATCAAGAGCGCGATGATGATGATGATGATGATGATGATGATGAT	1929						

QY	252	GRACAAACACTTCAGAGCTGCGGCGCCCCCAATCCCGGCTTATGCGAGAGTAT	311
Db	1	GTACAAACACTTCAGAGCGCGCGCCCCCGCCCCCGGCTTATGCGAGAGTAT	60
QY	312	GTATCAAGAGCCAGTGTATGGCAATTAATTTGGTCAAGAGGTGGTTACGCTGATACCGCTA	371
Db	61	TTACAGAGATGATTTTATG---TGCAGACATTTATGGTGTATGCTGACATACCGCTA	117
QY	372	CGCCCAAGCCCACTCCCTGCGCACTGCTGCTGCTCAAGTACAGTACGAGATTATGC	431
Db	118	CGCCCAAGCTTACCCCTGCGCACTGCGCTGCTCAAGTACAGTACGAGATTATGCT	177
QY	432	TGCGGACCCCTACACACACACACTTGTCTGCTGAGCCCCCACTACGCGGCTGGTCCATGAA	491
Db	178	TGCGGACCCCTACACACACACACTTGTCTGAGCCCCCACTACGCGGCTGGTCCATGAA	237
QY	492	TGCTTTTGGCGCCCTTGACCGCATGTCGCAAGACTAGAGCCATGCTGATGATGGTCTCGT	551
Db	238	TGCTTTTGGCACTTTGACCTGATGTCGCAAGACTAGAGCCATGCTGATGATGGTCTCGT	297
QY	552	TTCTTTCTTCATTCGACGAGCTAGTATATATACCAAGGGGGATCAACCGTTTGTCTCCATA	608
Db	298	TTCTTTCTTCATTCGACGAGCTAGTATATATACCAAGGGGGATCAACCGTTTGTCTCCATA	354
RESULT 14			
LOCUS	A1968176	439 bp	mRNA linear EST 25-AUG-1999
DEFINITION	wu14a06.x1 NCI CGAP GC6 Homo sapiens cDNA clone IMAGE:2516914 3'		
	similar to TR:043251 043251 HYPOTHETICAL 39.5 KD PROTEIN. ; mRNA		
ACCESSION	A1968176		
VERSION	A1968176.1	GI:5764994	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 439)		
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
JOURNAL	Tumor Gene Index		
COMMENT	Unpublished (1997)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgsbpa-remail.nih.gov		
	Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael		
	R. Emmer-Buck, M.D., Ph.D.		
	cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima		
	Bonaldo, Ph.D.		
	cDNA Library Arrayed by: Greg Lennon, Ph.D.		
	DNA Sequencing by: Washington University Genome Sequencing Center		
	Clone distribution: NCI-CGAP clone distribution information can be		
	found through the I.M.A.G.E. Consortium/BLND at:		
	www-bio.lnli.gov/bbbrp/image/image.html		
	Seq primer: -40UP from Gibco.		
FEATURES			
SOURCE			
	Location/Qualifiers		
	1..439		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:2516914"		
	/clone_1ib="NCI-CGAP_GC6"		
	/tissue_type="pooled germ cell tumors"		
	/lab_host="DH10B"		
	/note="Vector: pRTT3D-Pac (Pharmacia) with a modified		
	polylinker. Site 1: Not I; Site 2: Eco RI; Plasmid DNA		
	from the normalized library NCI-CGAP GC4 was prepared, and		
	se circies were made in vitro. Following HAP purification,		
	this DNA was used as tracer in a subtractive hybridization		
	reaction. The driver was PCR-amplified cDNAs from a pool		
	of 5,000 clones made from the same library (clonoids		
	1257096-1258631, 1469064-1470983, and 1475592-1476743).		
	Subtraction by Bento Soares and M. Fatima Bonaldo."		
BASE COUNT	91 a 142 c 114 g 92 t		

ORIGIN	Query Match	45.5%	Score 277.6	DB 9	Length 439
	Beet Local Similarity	84.7%	Pred. No. 2.3e-71		
	Matches 342	Conservative	0	Mismatches 19	Indels 43
					Gaps 1
Oy	205	GCAGTGCATATCCAGAGGAGCTCATCTTCGAGAGCGGTGTCACCGGTATCAACACCTTC	264		
Db	1	CGGCGCCCTTACCGAGGGGCGCACTTCGAGAGCGCGGTGTCACCGGTATCAACACCTTC	60		
Oy	265	AGAGCTGCGGCGCCGCCACCCCAATCCCGGCTATGCGGAGTAGTGTATCAAGAGCA	324		
Db	61	AGGGCGGCGGCGCCCGCCCGCCGATCCGGCTTACGCG-----	100		
Oy	325	GTGATGCAATTAATTGCTACAGGGTGTATACCTGTGATCCGCTACGCCACCCACC	384		
Db	101	-----GGTGTATATCTGCTATCCGCTACGCCACCTTACC	137		
Oy	385	CCTGCACTGCTGCTGCTGCTACAGTACAGATGAGAGAGATTTATGCTGCGGACCCCTAC	444		
Db	138	CTGTCACATGCGCTGCTGCTTACAGTACAGATTTATGCTGCGGACCCCTAC	197		
Oy	445	CACCACACACTTGTCTCAGAGCCCGGCTGAGCGGTGTCATGAATGCTTTTGCGCC	504		
Db	198	CACCACGCACTTGTCTCAGAGCCCGGCTGAGCGGTGTCATGAATGCTTTTGCACT	257		
Oy	505	TTGACCGATGCCAAGACTAGAGCCATGCTGATATGTTGGTCTCTTCTTCAATTG	564		
Db	258	TTGACTGATGCGAAGACTAGAGCCATGCTGATGATGTTGGTCTCTTCTTCAATTG	317		
Oy	565	CAGGCTAGTATATCAAGAGGGAGATACACCGTTTGTGCTCCATA	608		
Db	318	CAGGCTAGTATATCAAGAGGGAGATACACCGTTTGTGCTCCATA	361		
RESULT 15					
LOCUS	AMS89795	439 bp	MRNA	linear	EST 22-MAR-2000
DEFINITION	Hg2204.4.X1 NCI CGAP GC6 Homo sapiens cDNA clone IMAGE:2946343 3'				
	similar to TR:Q43251 Q43251 HYPOTHETICAL J9.5 KD PROTEIN. ;, mRNA				
	sequence.				
ACCESSION	AMS89795				
VERSION	AMS89795.1	GI:7276916			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 439)				
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),				
	Tumor Gene Index				
JOURNAL	Unpublished (1997)				
COMMENT	Contact: Robert Strausberg, Ph. D.				
	Email: cgaps-remail.nih.gov				
	Tissue Procurement: Christopher A. Moskalko, M.D., Ph.D., Michael				
	R. Emmert-Buck, M.D., Ph.D.				
	cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima				
	Bonaldo, Ph.D.				
	cDNA Library Arrayed by: Greg Lennon, Ph.D.				
	DNA Sequencing by: Washington University Genome Sequencing Center				
	Clone distribution: NCI-CGAP clone distribution information can be				
	found through the I.M.A.G.E. Consortium/LNL at:				
	image.llnl.gov/image/html/iresources.shtml				
FEATURES					
source	1. 439				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:2946343"				

```
QY 523 AGAGCCATGCTGATGATGATGGGCTCTGTTCTTTCTTCTTCAAGGCTAGATATATACCA 582
DB 298 AGAGCCATGCTGATGATGATGGGCTCTGTTCTTTCTTCTTCAAGGCTAGATATATACCA 357
QY 583 GGGGATACACCGCTTTGCTCCATA 608
DB 358 GGGGATACACCGCTTTGCTCCATA 383

RESULT 12
AUI45000/c 607 bp mRNA linear EST 05-AUG-2002
LOCUS AUI45000 HEMBA1 Homo sapiens CDNA clone HEMBA1003591 3', mRNA
DEFINITION AUI45000 HEMBA1 Homo sapiens CDNA clone HEMBA1003591 3', mRNA
sequence.
ACCESSION AUI45000
VERSION AUI45000.1 GI:11006521
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 607)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saio,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura,Y., Nagai,T., Sugano
,S. and Isogai,T.).
HRI human CDNA project (Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S.,
Saio,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura
,Y., Nagai,T., Sugano,S., Isogai,T.)
UNPUBLISHED (2000)
JOURNAL Contact: Takao Isogai
COMMENT Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@n1.co.jp
HRI human CDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; CDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source location/Qualifiers
1..607
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1003591"
/clone_lib="HEMBA1"
/tissue_type="whole embryo, mainly head"
/dev_stage="embryo, 10 weeks"
/vec="vector: pME18SFL3"
BASE COUNT 140 a 152 c 179 g 129 t 7 others
ORIGIN
Query Match 49.2%; Score 299.8; DB 9; Length 607;
Best Local Similarity 79.1%; Pred. No. 7e-78;
Matches 389; Conservative 0; Mismatches 99; Indels 4; Gaps 3;
```

```
DB 355 TCGTCATACCGCTACGCGCCAGCCCTACCCCTGCGCACTGCTTACAGTACAGTTA 296
QY 417 CGAGAGATTATATGCTGCGAGCCCTACACACACACTTCTCAGGCCCACTTACAGG 476
DB 295 CGAGAGATTATATGCTGCGAGCCCTACACACACACTTCTCAGGCCCACTTACAGG 236
QY 477 CGTTGGTCCATGAATGCTTTTGGCGCCCTTGACCGATGCCAAGACTAGAGCATGCTGA 536
DB 235 CGTTGGTCCATGAATGCTTTTGGCACTTTGACCTTGATGATGCCAAGACTAGAGCATGCTGA 176
QY 537 TGAATGGGCTCTGCTTTCTTTCTTCAAGGCTAGTATATACCAAGGGGATACACCG 596
DB 175 TGAATGGGCTCTGCTTTCTTTCTTCAAGGCTAGTATATACCAAGGGGATACACCG 116
QY 597 TTTTGTCTCCATA 608
DB 115 TTTTGTCTCCATA 104

RESULT 13
A1244212 459 bp mRNA linear EST 28-JAN-1999
LOCUS A1244212
DEFINITION q186b11.x1 NCI_CGAP_kid3 Homo sapiens CDNA clone IMAGE:1866333 3',
mRNA sequence.
ACCESSION A1244212
VERSION A1244212.1 GI:3839609
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 459)
AUTHORS NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgap@db-rcmail.nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNI, at:
www-bio.lnl.gov/bbrp/image/image.html
Insert Length: 452 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 451.
FEATURES
source location/Qualifiers
1..459
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1866333"
/clone_lib="NCI CGAP_Kid3"
/lab_host="DH10B"
/vec="vector: pTZ19-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand CDNA was primed with a Not I - oligo(dT) primer,
double-stranded CDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pTZ19 vector. mRNA
source: 2 pooled kidney. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaldo.
BASE COUNT 105 a 132 c 108 g 112 t 2 others
ORIGIN
Query Match 47.6%; Score 289.8; DB 9; Length 459;
Best Local Similarity 90.2%; Pred. No. 5.6e-75;
Matches 322; Conservative 0; Mismatches 32; Indels 3; Gaps 1;
```

ACCESSION A1374891
VERSION A1374891.1 GI:4174881
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 449)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert length: 400 Std Error: 0.00
Seq primer: -40UP from Gibco.
FEATURES
source
1. .449
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2048518"
/dev_stage="8-9 weeks"
/dev_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) focus material with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGCTTAATTTTCTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 93 a 136 c 114 g 105 t 1 others
ORIGIN
Query Match 51.6%; Score 314; DB 9; Length 449;
Best Local Similarity 90.2%; Pred. No. 3,6e-82;
Matches 348; Conservative 0; Mismatches 35; Indels 3; Gaps 1;
QY 223 GCTCACTTGAGGCGGTGTCGACCGGTGACCAACCTTCAGAGCTGCGCGCCCA 282
Db 1 GCGACCTGCGAGGCGCGGTGTCGACCGGTGACCAACCTTCAGAGCGCGCGCCCG 60
QY 283 CCCCCAATCCCGGCTATGCGGAGTAGTGTATCAAGACCGAGTATGCGCAATTAATG 342
Db 61 CCCCCGATCCCGGCTATGCGGAGTAGTGTATCAAGAGTAGATTTATAG--TGCAGAC 117
QY 343 CTACAGGTTGTTACGTCATACCGCTACGCGCCAGCCACCCCTGACAGCTGCTGCC 402
Db 118 ATTTATGTTGTTATGTCGATACCGCTACCGCCAGCTACCCCTGACAGCTGCTGCC 177
QY 403 TACAGTACAGTTACGAGCGAGTTATGCTGCGACCCCTACACACAGCACTTGCTCCA 462
Db 178 TACAGTACAGTTACGAGCGAGTTATGCTGCGACCCCTACACAGCACTTGCTCCA 237
QY 463 GCGCCCACTTACGCGGTGTCGATGAATGCTTTTGCGCCCTTACCGAGTCCAAAGCT 522
Db 238 GCGCCCACTTACGCGGTGTCGATGAATGCTTTTGCGACCTTTGACATGATGCCAAAGCT 297
QY 523 AGAGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 582
Db 298 AGAGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 357
QY 583 GGGGATACAAACCTTTTGCTCATA 608
Db 358 GGGGATACAAACCTTTTGCTCATA 383
RESULT 11

A1917298
LOCUS A1917298 456 bp mRNA linear EST 14-DEC-1999
DEFINITION ts96g09.x1 NCI_CGAP_G66 Homo sapiens cDNA clone IMAGE:2239168 3',
mRNA sequence.
ACCESSION A1917298
VERSION A1917298.1 GI:5637153
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 456)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher A. Moskalko, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/lml.at:
www-bio.lnl.gov/bdrip/image/image.html
Insert length: 478 Std Error: 0.00
Seq primer: -40UP from Gibco.
FEATURES
source
1. .456
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2239168"
/clone_lib="NCI CGAP GC6"
/issue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA
from the normalized library NCI CGAP GC4 was prepared, and
as circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIds
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 97 a 137 c 115 g 107 t
ORIGIN
Query Match 51.6%; Score 314; DB 9; Length 456;
Best Local Similarity 90.2%; Pred. No. 3,7e-82;
Matches 348; Conservative 0; Mismatches 35; Indels 3; Gaps 1;
QY 223 GCTCACTTGAGGCGGTGTCGACCGGTGACCAACCTTCAGAGCTGCGCGCCCA 282
Db 1 GCGACCTGCGAGGCGCGGTGTCGACCGGTGACCAACCTTCAGAGCGCGCGCCCG 60
QY 283 CCCCCAATCCCGGCTATGCGGAGTAGTGTATCAAGACCGAGTATGCGCAATTAATG 342
Db 61 CCCCCGATCCCGGCTATGCGGAGTAGTGTATCAAGAGTAGATTTATAG--TGCAGAC 117
QY 343 CTACAGGTTGTTACGTCATACCGCTACGCGCCAGCCACCCCTGACAGCTGCTGCC 402
Db 118 ATTTATGTTGTTATGTCGATACCGCTACCGCCAGCTACCCCTGACAGCTGCTGCC 177
QY 403 TACAGTACAGTTACGAGCGAGTTATGCTGCGACCCCTACACACAGCACTTGCTCCA 462
Db 178 TACAGTACAGTTACGAGCGAGTTATGCTGCGACCCCTACACAGCACTTGCTCCA 237
QY 463 GCGCCCACTTACGCGGTGTCGATGAATGCTTTTGCGCCCTTACCGAGTCCAAAGCT 522
Db 238 GCGCCCACTTACGCGGTGTCGATGAATGCTTTTGCGACCTTTGACATGATGCCAAAGCT 297

BASE COUNT 184 a 232 c 185 g 170 t 2 others
ORIGIN
Query Match 55.6%; Score 338.8; DB 10; Length 773;
Best Local Similarity 88.4%; Pred. No. 2,1e-89;
Matches 380; Conservative 0; Mismatches 47; Indels 3; Gaps 1;
Qy 179 GCTTTCATATCCGGCCCACTGCTGAGTGCATACCGAGGGCTCACTTCGAGGCC 238
Db 1 GCTTCCCGTATCCAGCAGCCACCGTCGCGCGCTTACCGAGGGCGACCTCGAGGCC 60
Qy 239 GTGCTCCGACCGTGACAAACACTTCAGAGTCGCGCGCCCAACCCCAATCCCGGCT 298
Db 61 GCGGTCCGACCGTGACAAACACTTCAGAGCGCGCGCCCGCCCGCGATCCGCGCT 120
Qy 299 ATGCGGAGTATGATATCAAGAGCCAGTGTATGGCAATAATGCTACAGGGTGTACG 358
Db 121 ACGGCGGTGTCTTACCAAGATGATTTATG---TGACAGACTTTATGTTGTTATG 177
Qy 359 CTGCATACCGCTAGCGCCGACCCCTGCGCACTGCTGCTTACAGTGAAGTACG 418
Db 178 CTGCATACCGCTAGCGCCGACCTTACCTGCGCACTGCGCTTACAGTGAAGTACG 237
Qy 419 GACGAGTTATGCTGCGCAGCCCTTACCAACACTTGTCTTCAAGCCCACTTACGGCG 478
Db 238 GACGAGTTATGCTGCGCAGCCCTTACCAACACTTGTCTTCAAGCCCACTTACGGCG 297
Qy 479 TTGGTGCATATGCTTTTGGGCGCTTGCAGCGATGCCAAGACTGAGAGCCATGCTGATG 538
Db 298 TTGGTGCATATATGCTTTTGGGCGCTTGCAGCGATGCCAAGACTGAGAGCCATGCTGATG 357
Qy 539 ATGTGGTCTGCTTTCTTCTTATTCAGAGCTAGTATATACCAAGGGGAGTACAACCGTT 598
Db 358 ATGTGGGCTCTGCTTTCTTCTTATTCAGAGCTAGTATATACCAAGGGGAGTACAACCGGT 417
Qy 599 TTGCTCCATA 608
Db 418 TTGCTCCATA 427
RESULT 7
AM197589 477 bp mRNA linear EST 29-NOV-1999
LOCUS xM44902.x1 NCI CGAP GC6 Homo sapiens cDNA clone IMAGE:2687090 3'
DEFINITION similar to TR:Q43251 Q43251 HYPOTHETICAL 39.5 KD PROTEIN.; mRNA
sequence.
ACCESSION AM197589
VERSION AM197589.1 GI:6476819
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 477)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LML at:
www-bio.lml.gov/bbrp/image/image.html
Possible reversed clone: similarity on wrong strand
Possible reversed clone: polyT not found
Seq primer: -40UP from Gibco.

FEATURES
source
Location/Qualifiers
1..477
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2687090"
/clone_1lb="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA
from the normalized library NCI CGAP GC4 was prepared, and
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneids
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 100 a 145 c 123 g 109 t
ORIGIN
Query Match 53.7%; Score 327.2; DB 10; Length 477;
Best Local Similarity 89.9%; Pred. No. 4.6e-86;
Matches 363; Conservative 0; Mismatches 38; Indels 3; Gaps 1;
Qy 205 GCAGCTCATACCGAGGGCTCACTTCGAGGCCGTGTGCGCACCGTGTACACACCTTC 264
Db 1 GCGGCGGCTTACCGAGGGCGCACCTTCGAGGCCGTGTGCGCACCGTGTACACACCTTC 60
Qy 265 AGAGCTGGGGGGCCCCCAACCCGCAATCCGGCTATGGCGGAGTATCAAGAGCCA 324
Db 61 AGGCGCGGGGGCCCCCGCCCGATCCGGCTACGGCGGAGTGTATACAGAGATGGA 120
Qy 325 GTGTATGGCAATAATTGCTACAGAGGTGTTACGCTGATACCGCTACGCCAGCCCAAC 384
Db 121 TTTATG---TGACAGACTTTATGTTGTTATGCTGATACCGCTACGCCAGCTTACC 177
Qy 385 CTGCGCACTGCTGCTGCTTACAGTGAAGTATCGAGAGATTATGCTGCCGACCCCTAC 444
Db 178 CTTGCGCACTGCTGCTGCTTACAGTGAAGTATCGAGAGATTATGCTGCCGACCCCTAC 237
Qy 445 CACCAACACTTGTCTCAGAGCCCGCACTAGCGCGTGTGCGAGTATGCTTTGGCGCC 504
Db 238 CACCAAGACTTGTCTCAGAGCCCGCACTAGCGCGTGTGCGAGTATGCTTTGGCACTT 297
Qy 505 TTGACCATGCCAAGATAGAGAGCCATGATGATGTGGTCTGCTTCTTCTTCAATTG 564
Db 298 TTGACTATGTCGAAGACTAGAGAGCCATGATGATGTGGTCTGCTTCTTCTTCAATTG 357
Qy 565 CAGGCTAGTATATCAAGAGGGATACACCGTTTGTCTCCATA 608
Db 358 CAGGCTAGTATATCAAGAGGGGATACACCGTTTGTCTCCATA 401
RESULT 8
AI401040 480 bp mRNA linear EST 30-MAR-1999
LOCUS th27a12.x1 NCI CGAP P-28 Homo sapiens cDNA clone IMAGE:2119486 3',
DEFINITION mRNA sequence.
ACCESSION AI401040
VERSION AI401040.1 GI:4244127
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 480)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.

Best Local Similarity 94.6%; Pred. No. 4.5e-95;
Matches 370; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY	205	GCAGTGCATACCGAAGGGGCTCACCCTTCAGAGGCCGTGTCGACACGTGTACAAACACTTC	264
Db	1	GGGCGCCCTACCGAGGGGCGACCTGCGAGGCCGGGTGCGACCGGTGTACAAACACTTC	60
QY	265	AGAGCTGGCGGCCCCCCACCCCAATCCCGGCTTATGCGGAGTACTGTATCAAGAGCA	324
Db	61	AGGGCCGGCGGCCCCCGCCCGCCGATCCCGGCTTACGGCGAGTACTGTATCAAGAGCT	120
QY	325	GTGTATGGCAATTAATTGCTATCAGGGTGGTTACGCTGCATACCGCCTACGCCCCAGCCACC	384
Db	121	GTGTATGGCAATTAATTGCTATCAGGGTGGTTATGCTGCATACCGCTACGCCCCAGCCCTACC	180
QY	385	CCTGCACCTGCTGCCTGCCTACAGTGCAGATTAACGACAGATTATGCTGCCGACCCCTAC	444
Db	181	CCTGCACCTGCGCGCTGCCTACAGTGCAGATTAACGACAGATTATGCTGCCGACCCCTAC	240
QY	445	CACCAACACATTGCTTCAGCCCCCACCCTTACGGGCTTGTCACATGATCTTTGGCGCC	504
Db	241	CACCAACGACCTTGCTTCAGCCCCCACCCTTACGGGCTTGTCACATGATCTTTGCACTT	300
QY	505	TTGACCGATGCCCAAGACTAGAGCCATGCTGATGATGTGGGTCTGTTCTTTTATTATG	564
Db	301	TTGACTGATGCCCAAGCTAGAGCCATGCTGATGATGTGGGTCTGTTCTTTTATTATTG	360
QY	565	CAGGCTAGTATATACCAAGGGGATACAAAC	595
Db	361	CAGGCTAGTATATACCAAGGGGATACAAAC	391

RESULT 5	AV729057	774 bp	mRNA	linear	EST 17-OCT-2000
LOCUS	AV729057				
DEFINITION	AV729057	HTC Homo sapiens	CDNA	clone HTCBG01 5',	mRNA sequence.
ACCESSION	AV729057				
VERSION	AV729057.1	GI:10838478			
KEYWORDS	EST.				
SOURCE	human.				

ORGANISM	REFERENCE
Homo sapiens	Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, G., Xiao, H., Xu, X.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	1 (bases 1 to 774)

TITLE	Human Sapiens CDNA HTC clones
JOURNAL	Unpublished (2000)
COMMENT	Contact: Zeguang Han

Chinese National Human genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong,
201203, P. R. China
Tel: 86-21-50801919 (ex. 45)
Fax: 86-21-50801922
Email: hanage@chc.sh.cn
This clone is available at CHGC in Shanghai.

```

FEATURES
    source
        location/Qualifiers
            1..774
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="HTC8G01"
                /clone_1fb="HTC"
                /tissue_type="Hypothalamus"
                /dev_stage="Adult"
                /lab_host="SOLR"
                /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

```

BASE COUNT	186 a	217 c	180 g	191 t
ORIGIN				
Query Match	56.2%	Score 342	DB 10	Length 774

Best Local Similarity -88.8%; Pred.No.2.4e-90;
Matches 382; Conservative 0; Mismatches 45; Indels 3; Gaps 1;

Oy	179	GCTTTCGATATCCGGCCCGCACACTGCTGCAGCTGCATACCGAGGGGCTCACTTCGAGGCC	238
Db	1	GCTTCCCGTATCCAGCACCACACGTCGCGGGCGCTTACGAGGGGCGCACCTCGCAGGCC	60
Oy	239	GTGTCGCAACCGGTGTACAACAACCTTCAGAGCTGCGGGCCCCCACCCCAATCCGGGCT	298
Db	61	GCGGTCGACCGGTGTACAACAACCTTCAGGGGCGGGCCCCCGCCCGATCCGGGGCT	120
Oy	299	ATGGCGGATGTGTGTATACAGAGCCAGTGTATGGCAATAAATTGCTACAGGTCGTTACG	358
Db	121	ACGGCGGTGTGTATTACAGAGTGAATTTTATGG---TGCAGACATTTATGTGTTATG	177
Oy	359	CTGCATACCGGTACGCGCCAGCCACCCCTGCACCTGTGCTGTGCTTACAGTGCAGTTACG	418
Db	178	CTGCATACCGGTACGCGCCAGCCCTGCACACTGCCTGCACCTGCCTTACAGTGCAGTTACG	237
Oy	419	GACGATTTATGCTGCGGCAACCCCTACACACACACTGTGCTCCAGCCCCCACTAGGGG	478
Db	238	GACGATTTATGCTGCGGCAACCCCTACACACACACTGTGCTCCAGCCCCCACTAGGGG	297
Oy	479	TTGCGCATGATATCTTTTGCGCCCTTGACCGATGCCAAGACTAGAGCCATGCTGATG	538
Db	298	TTGCGCATGATATCTTTTGACCTTTTGACCTTTGACGTAGGCCAAGACTAGGCAATGCTGATG	357
Oy	539	ATGTGGGTCGTGTTCTTTCTTCATTTGCAGGCTAGTATATCCAGGGGGATACAAACGTT	598
Db	358	ATGTGGGTCGTGTTCTTTCTTCATTTGCAGGCTAGTATATACGAGGGGGATACAAACGTT	417
Oy	599	TTGCTCCCAT 608	
Db	418	TTGCTCCCAT 427	

RESULT 6	AV729198	LOCUS	DEFINITION
AV729198	773 bp	mRNA	linear
AV729198	HTC Homo sapiens	CDNA clone	HTCBCD09 5', mRNA sequence.

ACCESSION VERSION KEYWORDS	AV729198	
SOURCE	EST.	
ORGANISM	human.	
	Homo sapiens	
		GI : 10838619

REFERENCE
AUTHORS

1 (bases 1 to 773)
Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, G., Xiao, H., Xu, X., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R.,

TITLE Chen, W., Chen, Z. and Han, Z.
JOURNAL Homo sapiens cdNA HTC clones
COMMENT Unpublished (2000)
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China

```

FEATURES
    source
        1. 773
        /Organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="HTCBCD09"
        /clone_lib="HTC"
        /tissue_type="Hyoorthalamus"

```

```

/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pbluescript sk(-); Site_1: EcORI; Site_2:
XhoI"

```

Db	236	CTACAGTGCAGTTAAGGACGAGTTATGCTCCGACCCTTAACAACACACTTGCTTC	177
OY	462	AGCCCCCACTTAGCGCGTTGGTGCCATGAAATGCTTTGGCCCTTGACCGATGCCAAGAC	521
Db	176	AGCCCCCAATCATGGCGGTGGTGCCATGAAATGCTTTGGCCCTTGACCGAAGCAAGAC	117
OY	522	TAGAGCGCATGTGATNGATNGTGGGTCTGCTGTTCTTCTTCATTGCAAGGCAATATATACCA	581
Db	116	TAGAGCGCATGTGATNGATNGTGGGTCTGCTGTTCTTCTTCATTGCAAGGCTAATATATACCG	57
OY	582	AGGGGATACCAACCGTTTTGCTCATAAT	609
Db	56	AGGGGATACCAACCGTTTTGCTCATAAT	29
RESULT 3			
LOCUS	A1656926	495 bp	mRNA linear EST 04-MAY-1999
DEFINITION	ct48e08.x1 NCI CGAP GC6 Homo sapiens cDNA clone IMAGE:2244038 3'		similar to TR:O43251 O43251 HYPOTHETICAL 39.5 KD PROTEIN.; mRNA sequence.
ACCESSION	A1656926		
VERSION	A1656926.1	GI:4740905	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: rsgabbs@emall.nih.gov Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/UM/LNL at: www.bio.linnl.gov/bdrip/image/image.html Seq primer: -40UP from Gdbco High quality sequence stop: 487. location/Qualifiers		
FEATURES			
Source	1. 495		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:2244038"		
	/clone_lib="NCI CGAP GC6"		
	/tissue_type="pooled germ cell tumors"		
	/lab_host="DH10B"		
	/note="Vector: p7T7D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; plasmid DNA from the normalized library NCI CGAP GC4 was prepared, and 88 circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."		
BASE COUNT	109 a 150 c 128 g 107 t		1 others
ORIGIN			
Query Match	60.8%; Score 370.4; DB 9; Length 495;		
Best Local Similarity	94.8%; Pred. No. 7.1e-99;		
Matches 383; Conservative	0; Mismatches 21; Indels 0; Gaps 0;		
OY	205	GCAGCTGCATACCGAGGGCTCACCTTCGAGGCCGTGGTCGACCGGTACAAACCTTC	264

Db	1	GGGGCCGGCTTACCGAGGGGGCGACCTGTGGAGGCGCGGGTCCGACCGGTATACAAACACTTC	60
Oy	265	AGAGTGTGGGGGGCCCCCAACCCCCCAATCCCCGGCTTATGGCGGAGTATATCAAGAGCA	324
Db	61	AGGGCCGGGGGGCCCCCGCCCCCGAGTCCCGGCTTACGGGAGTAGTATCAAGAGCTT	120
Oy	325	GTTGTATGCAATTAATTTCTCTACAGAGGTGGTTACGCTGATACCGGTATCGCCAGCCACC	384
Db	121	GTGTATGGCAATTAATTTCTCTCAGAGGTGGTTATGTGTGATACCGGTACGCCAGCTTACC	180
Oy	385	CTGTGCACTGTGTGTGCTTACAGTACAGTACAGGAGTTTATGTCTGCGGACCCCTTAC	444
Db	181	CTGTGCACTGTGCGGTGCTTACAGTACAGTACAGGAGTTTATGTCTGCGGACCCCTTAC	240
Oy	445	CACCAACACTTTGCTTCCAGCCGCCCACTTACAGGCGGTGGTGGCCATGAATGTTTGGGCC	504
Db	241	CACCAACGACTTGTCTCCAGCCGCCCACTTACAGGCGGTGGTGGCCATGAATGTTTGGCACTT	300
Oy	505	TTGACCGATGCGAAGACTAGGAGGCGCATGCTGTATGTATGGGCTGTCTTCTTCTCATTTG	564
Db	301	TTGACTGTATGCGAAGACTAGGAGGCGCATGCTGTATGTATGGGCTGTCTTCTTCTCATTTG	360
Oy	565	CAGGCTAGTATATATCAAGAGGGGGATACACACCGTTTGTCTCCATA	608
Db	361	CAGGCTAGTATATATCAAGAGGGGGATACACACCGTTTGTCTCCATA	404
RESULT 4			
LOCUS	A1186273	391 bp	mRNA linear EST 28-OCT-1998
DEFINITION	gd20b08.x1 Soares Placenta 80c9weeks 2bDHPel09W Homo sapiens cDNA		
VERSION	A1186273		
KEYWORDS	A1186273.1 GI:3736911		
SOURCE	EST.		
ORGANISM	human.		
	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	1 (bases 1 to 391)		
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .		
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
	Tumor Gene Index		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-rcmail.nih.gov This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 1800 Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence stop: 369. Location/Qualifiers		
FEATURES			
source	1..391		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:1724247"		
	/cdone_id="Soares_Placenta_80c9weeks_2bDHPel09W"		
	/clon_stage="Two placentae: one from 8 weeks and another		
	from 9 weeks post conception"		
	/lab_host="DH10B (ampicillin resistant)"		
	/note="Organ: placenta; Vector: pT773D (Pharmacia) with a		
	modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st		
	strand cDNA was primed with a Not I - oligo(dT) primer [5'		
	TGTTACCAATGTAGAGGAGGCGCGCGCATATTTTATTTTATTTT 3']		
	double-stranded cDNA was size selected, ligated to Eco RI		
	adapters (Pharmacia), digested with Not I and cloned into		
	the Not I and Eco RI sites of a modified pT773 vector		
	(Pharmacia). Library constructed by Bento Soares and		
	M.Fatima Bonaldo."		
BASE COUNT	73 a 129 c 107 g 82 t		
ORIGIN			
Query Match	58.7%; Score 357.4; DB 9; Length 391;		

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 32 Row: 1 Column: 9
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 8922072
This clone has the following problem: frame shifted.

FEATURES

source
1. 4039
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4815500"
/tissue_type="brain, hippocampus"
/clone_lib="NIH MGC_95"
/lab_host="DH10B"
/note="Vector: pBluescript"
BASE COUNT 1201 a 858 c 876 g 1104 t
ORIGIN

Query Match 80.0%; Score 487.2; DB 11; Length 4039;
Best Local Similarity 89.8%; Pred. No. 4,7e-133;
Matches 546; Conservative 0; Mismatches 58; Indels 4; Gaps 2;

```
OY 1 ATGACTAATAAAGGCGCTGAACCCCTACACCAATGGCTGAATTAATCCAGTTGG 60
DB 823 ATGACAAATAAAAAGCCCTCAACCTTTATCAAAATGGCTGAATTAATCAAGTTGG 882
OY 61 GCGCGGCTCTACAGCCCGCACTTTATGCAAGCAGCGTCTGTGTGCAAGCCAAACAG 120
DB 883 GGTGCGAGTCTACAGTCCCAATTTCTATGCAAGCAGCGTCTGTGTGCAAGCCAAACAG 942
OY 121 GAGGATCTTCATGATGAGTGGCCCACTTCACTTATATATCTTCAGAAAGCTGGC 180
DB 943 GAGGATCTTCATGATGAGTGGCCCACTTCACTTATATATCTTCAGAAAGCTGGC 1002
OY 181 TTTCATATCCGCGCGCAGCTGCTGAGCTGATACCAAGGGGCTCACTTCGAGGCGCT 240
DB 1003 TTCCCTATCCAGCAAGCCAGCGCGCGCCCTTACCAAGGGGCGAAGCTTCGAGGCGCG 1062
OY 241 GGTGCGACCGTGTACCAACCTTTCAGAGTGGCGGCGCCCGCCCAATCCCGGCTAT 300
DB 1063 GGTGCGACCGTGTACCAACCTTTCAGAGTGGCGGCGCCCGCCCGCAATCCCGGCTAT 1122
OY 301 GCGGAGTGTGTATGATGAGCCAGTGTATGCAATTAATGCTAGAGGTGTTACGCT 360
DB 1123 GCGGAGTGTGTATGATGAGCCAGATGATTTATAG--TGCAGACATTTATGCTATAGCT 1179
OY 361 GATATACCGCTACGCCAGCCAGCCAGCCCTGTCAGTGTGCTGCTACAGTACAGTTACGGA 420
DB 1180 GATATACCGCTACGCCAGCCAGCCAGCCCTGTCAGTGTGCTGCTACAGTACAGTTACGGA 1239
OY 421 CGAGTTATGCTGCGAGCCCTTACCAACCACTTGTCTTCAGCGCCCACTTACGCGGCT 480
DB 1240 CGAGTTATGCTGCGAGCCCTTACCAACCACTTGTCTTCAG--CCCACTTACGCGGCT 1298
OY 481 GGTGCGATGAATGCTTTTGGCGCTTGAAGCAAGTGAAGTGAAGTGAATGAT 540
DB 1299 GGTGCGATGAATGCTTTTGAAGCAAGTGAAGTGAAGTGAATGAT 1358
OY 541 GTGGGCTGCTTTCTTCTCATGAGGCTAGATATACCAAGGGGATACCAACGCTTTT 600
DB 1359 GTGGGCTGCTTTCTTCTCATGAGGCTAGATATACCAAGGGGATACCAACGCTTTT 1418
OY 601 GCTCATTA 608
DB 1419 GCTCATTA 1426
```

RESULT 2
BF549922/c 533 bp mRNA linear EST 12-DEC-2000
LOCUS BF549922

DEFINITION

UI-R-E0-bv-c-04-0-UI-r1 UI-R-E0 Rattus norvegicus cDNA clone

UI-R-E0-bv-c-04-0-UI 5', mRNA sequence.

BF549922

VERSION BF549922.1 GI:11659652

KEYWORDS EST.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE 1 (bases 1 to 533)

Bonaldi, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

9704447

JOURNAL MEDLINE

COMMENT

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

cDNA Library Preparation: M.B. Soares Lab Clone Distribution:

Clones will be available through Research Genetics (www.regen.com)

This clone is also available through the I.M.A.G.E. Consortium at

LLNL (info@image.llnl.gov). IMAGE ID= 1777166

Seq primer: M13 Forward.

Location/Qualifiers

1. 533

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-E0-bv-c-04-0-UI"

/dev_stage="embryonic"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pRT3D-Pac (Pharmacia) with a modified

polylinker. Site 1: NotI; Site 2: EcoRI; This library

consists of a mixture of individually tagged normalized

libraries constructed from 8, 12 and 18-day embryo. The

tag is a string of 3-5 nucleotides present between the

Not I site and the oligo-dT track which allows

identification of the library of origin of a clone within

the mixture."

BASE COUNT 134 a 125 c 164 g 109 t 1 others

ORIGIN

Query Match 69.5%; Score 423.4; DB 12; Length 533;

Best Local Similarity 92.7%; Pred. No. 1.4e-114;

Matches 471; Conservative 0; Mismatches 22; Indels 15; Gaps 2;

```
OY 114 CAACGAGGAGGATCTTCATATGACAGTGGCCCGGCTTCACTTGTATATCTTGGCAA- 172
DB 533 CAACGAGGAGGATCTTCATATGACAGTGGCCCGGCTTCACTTGTATATCTTGGCAA 474
OY 173 -----TGCGGCTTCCATATCCGCGCGCACTGCTGAGCTGATACGAGG 221
DB 473 TTCTGTCCAGTGCCTGCTGCTTTCATATCCGCGCGCACTGCTGAGCTGATACGAGG 414
OY 222 GGTCACTTTCAGGCGCGTGTGCGACCGGTATACACACTTTCAGAGCTGCGGCGCCCG 281
DB 413 GGTCACTTTCAGGCGCGTGTGCGACCGGTATACACACTTTCAGAGCTGCGGCGCCCG 354
OY 282 ACCCCCAATCCCGGCTTATGCGGAGTATGTATCAAGAGCCAGTATGCAATTAATT 341
DB 353 ACCCCCAATCCCGGCTTATGCGGAGTATGTATCAAGAGTATGCAATTAATT 297
OY 342 GGTACAGGTGTTACGCTGATACCGCTACCGCCAGCCCACTTCCAGCTGCTGCTGC 401
DB 296 CATTTATGATGATGCTGATACCGATACCGCCAGCCCACTTCCAGCTGCTGCTGC 237
OY 402 CTACAGTGAAGTTACGAGGATTAATGCTGCGACCCCTTACCAACACACTTGTCTCC 461
```

GenCore version 5.1.4.DS_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 12:40:33 ; Search time 856.8475 Seconds
(without alignments)
11515.871 Million cell updates/sec

Title: US-09-809-545A-1_COPY_535_1143

Perfect score: 609
Sequence: 1 atgactaataaaaaagccgcg.....acaaccgttcgctccatcat 609

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
EST :
1: em_gestba :
2: em_gesthum :
3: em_gestlin :
4: em_gestmu :
5: em_gestov :
6: em_gestpi :
7: em_gestto :
8: em_htc :
9: gb_est1 :
10: gb_est2 :
11: gb_htc :
12: gb_est3 :
13: gb_est4 :
14: gb_est5 :
15: em_estfun :
16: em_estom :
17: gb_gss :
18: em_gss_hum :
19: em_gss_inv :
20: em_gss_pln :
21: em_gss_vrc :
22: em_gss_fun :
23: em_gss_mam :
24: em_gss_mus :
25: em_gss_other :
26: em_gss_pro :
27: em_gss_rtd :

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	487.2	80.0	4039	11	BC026312 Homo sapi
2	423.4	69.5	533	12	BP549922 UI-R-B0-b
3	370.4	60.8	495	9	AI656926 t48e08.x
4	357.4	58.7	391	9	AI186273
5	342	56.2	774	10	AV729057
6	338.8	55.6	773	10	AV729198

7	327.2	53.7	477	10	AW197589	AW197589 xm44602.x
8	327.2	53.7	480	9	AI401040	AI401040 cb27a12.x
9	314	51.6	353	9	AI189407	AI189407 qd01h05.x
10	314	51.6	449	9	AI374891	AI374891 ca60g12.x
11	314	51.6	456	9	AI917298	AI917298 cb96g09.x
12	299.8	49.2	607	9	AUI45000	AUI45000 AUI45000
13	289.8	47.6	459	9	AI244212	AI244212 q186b11.x
14	277.6	45.6	439	9	AI968176	AI968176 wu14a06.x
15	277.6	45.6	439	10	AM589795	AM589795 hp22d04.x
16	277.6	45.6	439	10	BE501380	BE501380 7a41h09.x
17	275.8	45.3	521	13	BM183349	BM183349 fv10e03.x
18	275.8	45.3	723	12	BG306387	BG306387 fm58d11.x
19	269.6	44.3	367	9	AI095813	AI095813 gb20g11.x
20	261.8	43.0	420	9	AI655094	AI655094 wbs7a10.x
21	259.4	42.6	372	9	AI799929	AI799929 wca1a05.x
22	254.6	41.8	427	12	BF223478	BF223478 7q33e01.x
23	247.2	40.6	392	9	AA975235	AA975235 oq36c08.x
24	247.2	40.6	443	9	AA773715	AA773715 af81c05.x
25	228.6	37.5	680	11	AK005186	AK005186 Mus muscu
26	224	36.8	261	14	F21037	F21037 HSPD05E03.H
27	217.4	35.7	353	14	D60520	D60520 HUM114E06A
28	205.4	33.7	379	9	AI843387	AI843387 UI-M-AQ1-
29	198	32.5	268	9	AI175239	AI175239 EST218774
30	197.4	32.4	430	9	AI850237	AI850237 UI-M-BG1-
31	195.8	32.2	425	10	BB750925	BB750925 BB750925
32	191.2	31.4	218	9	AI189257	AI189257 qc39h03.x
33	187	30.7	327	12	BG186689	BG186689 RST5664.A
34	182.6	30.0	246	9	AI351484	AI351484 qf05e06.x
35	180.8	29.7	361	9	AI835011	AI835011 UI-M-AM1-
36	177.8	29.2	655	14	BO188308	BO188308 UI-E-EJ1-
37	177.4	29.1	449	9	AI335996	AI335996 qe42e01.x
38	176.8	29.0	451	9	AI193843	AI193843 qe72h11.x
39	175.2	28.8	347	9	AI041587	AI041587 ox56g05.x
40	175.2	28.8	479	9	AI291784	AI291784 qm75b08.x
41	174.2	28.6	283	12	BG185609	BG185609 RST4560.A
42	172.6	28.3	302	14	BM930547	BM930547 UI-E-EJ1-
43	172.6	28.3	973	11	BC033885	BC033885 Homo sapi
44	167.6	27.5	211	10	AV336519	AV336519 AV336519
45	166.8	27.4	457	9	AI193117	AI193117 qe69h03.x

ALIGNMENTS

RESULT 1
LOCUS BC026312 4039 bp mRNA linear HTC 08-APR-2002
DEFINITION Homo sapiens, clone IMAGE:4815500, mRNA.
ACCESSION BC026312
VERSION BC026312.1 GI:20070932
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (02-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550, USA

REMARK
COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: gcabs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiyaki Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LUNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-mgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu

THIS PAGE BLANK (USPTO)

	XN	Mus musculus.
	PX	DE20103510-U1.
	XX	
	PD	07-JUN-2001.
	XX	
	PF	28-FEB-2001; 2001DE-2003510.
	PR	02-DEC-1999; 99DE-1058160.
	PA	(LION-) LION BIOSCIENCE AG.
	DR	WPI; 2001-368570/39.
	PT	Gene library containing sequences with specific 3'-ends and no polyA tail, encoding proteins involved in a wide range of cellular processes
	PS	-
	XX	Claim 15; Page 174; 251pp; German.
	CC	This invention describes a novel gene library (A) comprises a gene sequence (or its part) encoding a protein involved in amino acid synthesis, cellular/energy metabolism, metabolism of fatty acids/phospholipids, synthesis or breakdown of purines/pyrimidines/nucleosides/nucleotides, DNA replication/transcription/translation, or is a transport/binding protein. (A) are produced that correspond to the 3'-end of mRNA but without the polyA tail. They can be prepared more efficiently and with less effort than conventional libraries. AAK53436-AAK54275 represent fragments of the gene library described in the method of the invention.
SQ	Sequence 327 BP; 71 A; 65 C; 84 G; 107 T; 0 other;	
	Query Match 21.9%; Score 133.6; DB 22; Length 327; Best Local Similarity 95.8%; Pred. NO. 8-2e-31; Matches 159; Conservative 0; Mismatches 4; Indels 3; Gaps 2	
OY	447 CCACACACTTGCTCCA--GCCCCCACCTTAGGGCGTTGTCGCATGAAGC-TTTTGGCCCC DB 323 CCACACACTGTCTCCAAAGCCCCCACCCCTACGGCGTTGGCATGAATCTTTTGGCGCC 264	
OY	504 CTTGACCATTGCCAACACTAGAAGCAATGCTATGATGTGGGCTTCGTTCTTCATTT 563 DB 263 CTGACCGATGCCAAGACTAGAGAGCCATGCTATGATGTGGGCTTCGTTCTTCATTT 204	
OY	564 GCAGCGTAGTATATATACAAGGGGGATATAACACGTTTTGCTCCATAT 609 DB 203 GCAGCGTAGTATATATACCAGGGGGATACAACCGTTTTGCTCCATAT 158	
RESULT 15		
AAS3229/c	ID AAKS3229 standard; cDNA: 1164 BP.	
XX AC	AAKS3229;	
XX XX		
DJ DT	06-NOV-2001 (first entry)	
XX XX		
DE DE	Human polynucleotide SEQ ID NO 2758.	
KM KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy; vaccine; peptide therapy; stem cell growth factor; haematopoiesis; tissue growth factor; immunomodulatory; cancer; leukaemia; nervous system disorder; arthritis; inflammation; ss.	
OS OS	Homo sapiens.	
XX PN	MOZ00157190-A2.	
PN PD	09-AUG-2001.	
XX PF	05-FEB-2001; 2001WO-US04098.	

```

XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0654936.
XX 15-SEP-2000; 2000US-0663561.
XX 20-OCT-2000; 2000US-0693325.
XX 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wehrman T, Goodrich R;
XX WPI: 2001-476283/51.
XX P-PSDB: AAM80096.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
XX useful in diagnosis and gene therapy -
XX
XX Claim 1; Page 4968; 6221p; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK51435) and the
XX encoded polypeptides (AAM78333-AAK80302) that exhibit activity relating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and the polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
XX (AAM80020) are omitted as the relevant pages from the sequence listing
XX were missing at the time of publication.
XX
XX Sequence 1164 BP; 273 A; 280 C; 325 G; 286 T; 0 other;
XX
XX Query Match 21.2%; Score 129; DB 22; Length 1164;
XX Best Local Similarity 90.2%; Prid. No. 3.ee-29;
XX Matches 138; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
XX
XX 1 ATGACTATAAAAGGCGCGTAGACCCCGACACCAATGCGGAATTAATCCAGTGTG 60
XX Db 212 ATGACAATAAAAGACCGCTCAACCCCTTATCAAAATGGCTGGAATTCAGTGTG 213
XX
XX 61 GGGCGGCTTACAGACCCCGACTTATGACAGGACGCGTGTGTGSCCAGGCCAACCAAG 120
XX Db 212 GGTGACAGCTACAGTCCCGAATTCCTTAGCAGGACGCGTCTGTGTGCCAGGCCAACCAAG 153
XX
XX 121 GAGGATCTTCAGTACAGTACAGTGGCCCAAGTTCA 153
XX Db 152 GAGGATCTTCAGTACAGTACAGTGGCCCAAGTTCA 120

```

Search completed: March 15, 2003, 12:51:58
Job time : 153.798 secs

OS	Mus musculus.
PN	DE20103510-U1.
PD	07-JUN-2001.
XX	
PF	28-FEB-2001; 2001DE-2003510.
XX	
PR	02-DEC-1999; 99DE-1058160.
PA	(LION-) LION BIOSCIENCE AG.
XX	
DR	WPI; 2001-368570/39.
PT	Gene library containing sequences with specific 3'-ends and no polyA tail, encoding proteins involved in a wide range of cellular processes
PS	Claim 15; Page 174; 251pp; German.
CC	This invention describes a novel gene library (A) comprises a gene sequence (or its part) encoding a protein involved in amino acid synthesis, cellular/energy metabolism, metabolism of fatty acids/phospholipids, synthesis or breakdown of purines/pyrimidines/nucleosides/nucleotides, DNA replication/transcription/translation, or is a transport/binding protein (A) are produced that correspond to the 3'-end of mRNA but without the polyA tail. They can be prepared more efficiently and with less effort than conventional libraries. AAK53436-AAK54275 represent fragments of the gene library described in the method of the invention.
SQ	Sequence 327 BP; 71 A; 65 C; 84 G; 107 T; 0 other;
Query Match	21.9%; Score 133.6; DB 22; Length 327;
Best Local Similarity	95.8%; Pred. No. 8.2e-31;
Matches 159; Conservative	0; Mismatches 4; Indels 3; Gaps
OY	447 CCACACACTTGCTCTCA--GCCCCCACCTAGCGGCGTTGGTGCATGAATGC-TTTGGGCC 503
Db	323 CCACACACTTGCTCCAAAGCCCCCACCCCTACGCGGTGGTCATGAATGCTTTTGGC 264
OY	504 CTTGACCGATCCCAAAGACTAGAGCCATGTCGATGANGTGGGCTCTGTTCTTCTCAT 563
Db	263 CTGACCGATGCCAAAGACTAGAGCATGTCGATGATGATGGGTCGTCGTTCTTCTCAT 204
OY	564 GCAGGCTGTATTATACCAAGGGGGATACAACCGTTTTGCTCATAT 609
Db	203 GCAGGCTGTATTATACCGAGGGGATACACCCTTTTGCTCATAT 158
RESULT 15	
AAK53229/c	
ID	AAK53229 standard; cDNA: 1164 BP.
XX	
AC	AAK53229;
XX	
DT	06-NOV-2001 (first entry)
XX	
DE	Human polynucleotide SEQ ID NO 2758.
KX	
KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy; vaccine; peptide therapy; stem cell growth factor; haematopoiesis; tissue growth factor; immunomodulatory; cancer; leukaemia; nervous system disorder; arthritis; inflammation; ss.
OS	Homo sapiens.
XX	
PN	WO200157190-A2.
PD	09-AUG-2001.
XX	
PF	05-FEB-2001; 2001WO-US04098.

CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 939 BP, 148 A, 248 C, 297 G, 246 T, 0 other;

Query Match 22.1%; Score 134.4; DB 22; Length 939;
Best Local Similarity 68.0%; Pred. No. 7.3e-31;
Matches 234; Conservative 0; Mismatches 101; Indels 9; Gaps 3;

QY 166 TCTGCAATGCTGGCTTTCATATCCGGCCGCACTGTGCAAGCTGATACCGAGGGCT 225
DB 417 TATGCAAGTACCGGGTCCCTTACCCACCGGCAAGCGGTTGCTTACCGGGCGCA 358
QY 226 CACCTTCAGGCGCGTGTGCGACCGGTATCAACCTTCAGAGCTGGCGGCCACCC 285
DB 357 CATCTTCGGGGGCGGGGCGGGGTATTAATCAATTCGGGCTGGCGCACCC 298
QY 286 CCAATCCCGGCTATGCGCGAGTATGATCAAGACCAAGTATGGAATATTGCTA 345
DB 297 CCCATCCGCACTTACGAGCGGTGCTATCAAGATGATTTATGCTGCTG---GATT 241
QY 346 CAGGCTGTACGCTGATACCGCTAGCCGACCCCACTGCTGCTGCTGCTAC 405
DB 240 TATGAGGCTACGAGCGCTACAGATAGCTCAGC---CCGCTGACGGCGGAGCTTAC 184
QY 406 AGTACAGTTAGGAGAGATTTA---TGCTGCCAGCCCTTACACACACACACTGCTCA 462
DB 183 AGCGACAGTTAGCGGAGAGCTTACGAGCTGCGGACCCGATCCATACCAATCGGGCC 124
QY 463 GCCCCACCTACGCGGCTGTGTCATGATGCTTTGGCCCTT 506
DB 123 GCGGCGACCTACAGCATTTGGAACATGTGAACCTTCCACCGTT 80

RESULT 13
AA159248
ID AA159248 standard; cDNA; 1011 BP.

AC AA159248;

XX 22-OCT-2001 (first entry)
DE Human polynucleotide SEQ ID NO 1451.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.

XX Homo sapiens.

OS WO200153312-A1.

PN 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX MPI; 2001-442253/47.
DR P-PSDB; AAM40092.

PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX Claim 1; SEQ ID NO 1451; 10078bp; English.

CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 1011 BP; 260 A; 314 C; 267 G; 170 T; 0 other;

Query Match 22.1%; Score 134.4; DB 22; Length 1011;
Best Local Similarity 68.0%; Pred. No. 7.5e-31;
Matches 234; Conservative 0; Mismatches 101; Indels 9; Gaps 3;

QY 166 TCTGCAATGCTGGCTTTCATATCCGGCCGCACTGTGCAAGCTGATACCGAGGGCT 225
DB 615 TATGCAAGTACCGGGTCCCTTACCCACCGGCAAGCGGTGCTTACCGGGCGCA 674
QY 226 CACCTTCAGGCGCGTGTGCGACCGGTATCAACACCTTCAGAGCTGGCGGCCACCC 285
DB 675 CATCTTCGGGGGCGGGGCGGGCGGTATTAATCAATTCGGGCTGGCACCCACCC 734
QY 286 CCAATCCCGGCTATGCGCGAGTATGATCAAGACCAAGTATGGAATATTGCTA 345
DB 735 CCCATCCGCACTTACGAGCGGTGCTGATCAAGATGATTTATGCTGCTG---GATT 791
QY 346 CAGGCTGTACGCTGATACCGCTACCGCCGACCCCTGCCACTGCTGCTTAC 405
DB 792 TATGAGGCTTACGAGCGCTTACAGATAGCTTACG---CCGCTGACGGCGGAGCTTAC 848
QY 406 AGTACAGTTAGGAGAGATTTA---TGCTGCCAGCCCTTACACACACACTGCTCA 462
DB 849 AGCGACAGTTAGCGGAGAGCTTACGAGCTGCGGACCCGATCCATACACATCGGGCC 908
QY 463 GCCCCACCTACGCGGCTGTGTCATGATGCTTTGGCCCTT 506
DB 909 GCGGCGACCTACAGCATTTGGAACATGTGAACCTTCCACCGTT 952

RESULT 14
AAK54001/C
ID AAK54001 standard; cDNA; 327 BP.

AC AAK54001;

XX 16-NOV-2001 (first entry)

DE Murine transcription associated protein encoding cDNA SEQ ID 566.

XX Murine; liver; gene library; amino acid synthesis; binding protein;
KW cell metabolism; energy metabolism; fatty acid metabolism; synthesis;
KW phospholipid metabolism; purine; pyrimidine; nucleoside; nucleotide;
KW replication; transcription; translation; transport protein; ss.

KM Human; nootropic; immunosuppressant; cyrostatic; gene therapy; Cancer;
 KM peripheral nervous system; neuropathy; central nervous system; CNS;
 KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KM chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KM leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI; 2001-442253/47.
 DR P-PSDB; AAM41878.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1; SEQ ID NO 5023; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cyrostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 SQ Sequence 939 BP; 148 A; 248 C; 297 G; 246 T; 0 other;
 XX
 XX
 Query Match 22.1%; Score 134.4; DB 22; Length 939;
 Best Local Similarity 66.0%; Pred. No. 7.3e-31;
 Matches 234; Conservative 0; Mismatches 10; Indels 9; Gaps 3;

QY 346 CAGGTGGTTACGCTGCTACCGCCAGCCACCCCTGCACCTGCTGCTTAC 405
 DB 240 TATGAGGAGCTACGAGCCTTACAGATACCTAGC---CCGCTGACGCGCGAGCCTTAC 184
 QY 406 AGTGACAGTTACGGACAGATTAA---TGCCTCCGACCCCTACACACACACTGCTCCA 462
 DB 183 AGGCACAGTTACGGACAGAGTCTACGCACTGCCAGCCGTCATCATACACATCGGGCC 124
 QY 463 GCCCCACCTACGGCGTGTGTCATGAAAGCTTTGCGCCCTT 506
 DB 123 GCGGCACTTACAGCATTGGAACCATGGAACCTTCCACCGTT 80
 RESULT 12
 AA161035/C
 ID AA161035 standard; cDNA; 939 BP.
 XX
 AC AA161035;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 5024.
 XX
 KM Human; nootropic; immunosuppressant; cyrostatic; gene therapy; Cancer;
 KM peripheral nervous system; neuropathy; central nervous system; CNS;
 KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KM chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KM leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI; 2001-442253/47.
 DR P-PSDB; AAM41879.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1; SEQ ID NO 5024; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cyrostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIP
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 567 BP; 92 A; 151 C; 169 G; 155 T; 0 other;
Query Match 22.1%; Score 134.4; DB 23; Length 567;
Best Local Similarity 68.0%; Pred. No. 5.9e-31;
Matches 234; Conservative 0; Mismatches 101; Indels 9; Gaps 3;
OY 166 TCTGGAATGCTGCTTCCATATCCGGCCGACCTGCTGCAAGCTGATACCGAGGGCT 225
DB 417 TATGACAGTACGGGGTCCCTTACCCACACCGGACAGCGGCTTACCGGGGCGCA 358
OY 226 CACCTTCGAGCGCGTGTGCGACCGGTGTAACAACCTTCAGAGCTGCGGCGCCACCC 285
DB 357 CATCTTCGGGGCGCGGGCGCGGCGGTATTAATCATTTGGGGCTGCGCCACCCACCC 298
OY 286 CCAATCCCGGCTATGCGCGAGTGTATCAAGAGCCAGTGTATGCAATTAATTGCTA 345
DB 297 CCAATCCCGGCTATGCGCGAGTGTATGATGATGATGATGATGATGATGATGATGAT 241
OY 346 CAGGCTGCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 405
DB 240 TATGAGAGCTTACGACCGCTTACAGATAGCTCAGCC--CCGCTGACGCGCGGAGCTAC 184
OY 406 AGTGACAGTTACGACGAGTTA---TGCTGCGACCCCTTACACACACACACTTGTCTCA 462
DB 183 ACGGACAGTTACGACGAGTTACGACGAGTGTGCGACCGCTTACATCATACACACTGCGGCC 124
OY 463 GCCCCCACTTACGCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 506
DB 123 GCGGCGACCTTACGACGAGTGTGAACATGTAACCTTCCACCGTT 80
RESULT 10
AA159249
ID AA159249 standard; cDNA; 918 BP.
XX
AC AA159249;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 1452.
XX
XX Human; nocotropic; immunosuppressant; cytosolic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
XX
PR 25-APR-2000; 2000US-0552317.
XX
PR 09-JUL-2000; 2000US-0598042.
XX
PR 19-JUL-2000; 2000US-0620312.
XX
PR 03-AUG-2000; 2000US-0653450.
XX
PR 14-SEP-2000; 2000US-0662191.
XX
PR 19-OCT-2000; 2000US-0693036.
XX
PR 29-NOV-2000; 2000US-0727344.

XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QH, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
DR P-PSDB; AAM40093.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
PS
PS Claim 1; SEQ ID NO 1452; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nocotropic,
CC immunosuppressant and cytosolic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 918 BP; 235 A; 300 C; 235 G; 148 T; 0 other;
Query Match 22.1%; Score 134.4; DB 22; Length 918;
Best Local Similarity 68.0%; Pred. No. 7.2e-31;
Matches 234; Conservative 0; Mismatches 101; Indels 9; Gaps 3;
OY 166 TCTGGAATGCTGCTTCCATATCCGGCCGACCTGCTGCAAGCTGATACCGAGGGCT 225
DB 522 TATGACAGTACGGGGTTCCTTACCCACACCGGACAGCGGCTTACCGGGGCGCA 581
OY 226 CACCTTCGAGCGCGTGTGCGACCGGTGTAACAACCTTCAGAGCTGCGGCGCCACCC 285
DB 582 CATCTTCGGGGCGCGGGCGCGGCGCGGTATTAATCATTTGGGGCTGCGCCACCCACCC 641
OY 286 CCAATCCCGGCTATGCGCGAGTGTATCAAGAGCCAGTGTATGCAATTAATTGCTA 345
DB 642 CCAATCCCGGCTATGCGCGAGTGTATGATGATGATGATGATGATGATGATGATGAT 698
OY 346 CAGGCTGCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 405
DB 699 TATGAGAGCTTACGACGAGTGTACAGATACCGTCAAGC---CCGCTGACGCGCGGACGCTAC 755
OY 406 AGTGACAGTTACGACGAGTTA---TGCTGCGACCCCTTACACACACACTTGTCTCA 462
DB 756 ACGGACAGTTACGACGAGTGTACGAGTGTGCGACCGCTTACATCATACACACTGCGGCC 815
OY 463 GCCCCCACTTACGCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 506
DB 816 GCGGCGACCTTACGACGAGTGTGAACATGTAACCTTCCACCGTT 859
RESULT 11
AA161034/C
ID AA161034 standard; cDNA; 939 BP.
XX
AC AA161034;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 5023.
XX

|||||
Db 241 CAGCCCCACCTACGGCGCTGGTGCATGAAGCTTTTGACCTTGAAGTATGCCAAGA 182
Qy 521 CTAGAGACCATGCTGATGATGAGTGTGCTCTCTTTTCATATGACAGCTATATACC 580
Db 181 CTAGAGACCATGCTGATGATGAGTGTGCTCTCTTTTCATATGACAGCTATATACC 122
Qy 581 AAGGGGATACACCGTTTGTCTCAT 608
Db 121 GAGGGGATACACCGTTTGTCTCAT 94
RESULT 8
AAS93632 standard; cDNA; 406 BP.
XX ID AAS93632 standard; cDNA; 406 BP.
XX AAS93632;
AC AAS93632;
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #29436.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM Food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX MO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001MO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR P-PSDB; ABG29445.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 29436; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 406 BP; 94 A; 97 C; 85 G; 102 T; 28 other;

Query Match 31.2%; Score 190; DB 23; Length 406;
Best Local Similarity 97.5%; Pred. No. 4,8e-48;
Matches 193; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 411 CAGTTAGGAGAGATTTATGCTGCGACCCCTACACACACACTTGTCCAGCCCCAC 470
Db 42 CAGTTAGGAGAGATTTATGCTGCGACCCCTACACACACACTTGTCCAGCCCCAC 101
Qy 471 CTAGAGCGTTGGTGCATGAATGCTTTTGCGCCCTTGACCGATCCAGATAGAGCCA 530
Db 102 CTAGAGCGTTGGTGCATGAATGCTTTTGCGACCTTTGACTGATGCCAAGATAGAGCCA 161
Qy 531 TGTGATGATGAGTGTGCTCTCTCTTTCTTTCATTGACAGCTATATATCAAGGGGGATA 590
Db 162 TGTGATGATGAGTGTGCTCTCTCTTTCTTTCATTGACAGCTATATATCAAGGGGGATA 221
Qy 591 CAACCGTTTGTCTCAT 608
Db 222 CAACCGTTTGTCTCAT 239
RESULT 9
AAS6245/C
XX ID AAS6245 standard; cDNA; 567 BP.
XX AAS6245;
AC AAS6245;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #2049.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM Food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX MO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001MO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR P-PSDB; ABG02058.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 2049; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in

AC AAS93633;
 DT 13-FEB-2002 (first entry)
 DE DNA encoding novel human diagnostic protein #29437.
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.
 PN WO200175067-A2.
 PD 11-OCT-2001.
 PF 30-MAR-2001; 2001WO-US08631.
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 PA (HXSE-) HXSEQ INC.
 PI Drmanac RT, Liu C, Tang YT;
 PI WPI: 2001-639362/73.
 PR P-PSDB: ABG29446.
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PS Claim 1; SEQ ID NO 29437; 103bp; English.
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XQ Sequence 481 BP; 100 A; 145 C; 123 G; 113 T; 0 other;

[illegible]

Dd	121	TTTTATGG---TGCAGACATTATGTTGTGTTTGCTGCATACCGCTACGCCAGCCTAAC	177
Oy	385	CCGCGCACCTGCTGCTGCTCAACAGTGAACAATTGACGACGAGTTTANG--CTGCCAGACCCT	442
Dd	178	CTGCGCACCTGCTGCTGCTCAACAGTGAACAATTGACGACGAGTTTANGACATTTTCCCCCT	237
Oy	443	ACGACACACACTTGTGCTCCAGCCCCCACCCTACGCGGTTGTGTCATGATGCTTTTGGCG	502
Dd	238	GTCACCACTACCTGCTCCAGCCCCCACCCTACGCGGTTGTGTCATGATGCTTTTGGAC	297
Oy	503	CTTTGACCGATGCGAAGACTGAGAGCCATGCTGATGATGTGGGTCGTTCTTTCTTCAT	562
Dd	298	CTTTGACCTGATGCGAAGACTGAGAGCCATGATGATGTGGGTCGTTCTTTCTTCAT	357
Oy	563	TGCAGGCTAGTATATATACCAAGGGCGATTAACCGTTTGGCTCCATA	608
Dd	358	TGCAGGCTAGTATATATACCGAGGGCGATTAACCGTTTGGCTCCATA	403
RESULT 6			
ID	AAH09205/c		
XX	AAH09205 standard; cDNA; 607 BP.		
XX	AAH09205;		
Dt	26-JUN-2001 (first entry)		
XX			
Dd	Human cDNA clone (3'-primer) SEQ ID NO:6040.		
XX			
KX	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.		
OS	Homo sapiens.		
XX	BPI074617-A2.		
Pn	07-FEB-2001.		
Pd			
XX	28-JUL-2000; 2000EP-0116126.		
Pf			
XX	29-JUL-1999; 99JP-0248036.		
PR	27-AUG-1999; 99JP-0300253.		
PR	11-JAN-2000; 2000JP-0118776.		
PR	02-MAY-2000; 2000JP-0183767.		
Pr	09-JUN-2000; 2000JP-0241899.		
XX			
PA	(HELI-) HELIX RES INST.		
Pi	Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
Pi	Iehii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
XI			
Dr	WPI; 2001-318749/34.		
XX			
Pt	Primer sets for synthesizing polynucleotides, particularly the 5602		
Pt	full-length cDNAs defined in the specification, and for the detection		
Pt	and/or diagnosis of the abnormality of the proteins encoded by the		
Pt	full-length cDNAs -		
PS			
Cc	Claim 3; SEQ ID 6040; 2537bp + CD ROM; English.		
Cc			
Cc	The present invention describes primer sets for synthesising 5602		
Cc	full-length cDNAs defined in the specification. Where a primer set		
Cc	comprises: (a) an oligo-dT primer and an oligonucleotide complementary		
Cc	to the complementary strand of a polynucleotide which comprises one of		
Cc	the 5602 nucleotide sequences defined in the specification, where the		
Cc	oligonucleotide comprises at least 15 nucleotides; or (b) a combination		
Cc	of an oligonucleotide comprising a sequence complementary to the		
Cc	complementary strand of a polynucleotide which comprises a 5'-end		
Cc	sequence and an oligonucleotide comprising a sequence complementary to a		
Cc	polynucleotide which comprises a 3'-end sequence, where the		
Cc	oligonucleotide comprises at least 15 nucleotides and the combination of		
Cc	the 5'-end sequence/3'-end sequence is selected from those defined in		
Cc	the specification. The primer sets can be used in antisense therapy and		
Cc	in gene therapy. The primers are useful for synthesising polynucleotides,		

```
Qy 481 GGTGCATGAATGCTTTTGGCCCTTGACCGATGCCAGACTAGAGCCATGCTGATGAT 540
Db 1570 GGTGCATGAATGCTTTTGGCCCTTGACCGATGCCAGACTAGAGCCATGCTGATGAT 1629
Qy 541 GTGGCTCTGCTTTCTTCTTCATTTGACGAGTACTATATATACCAAGGGGATACACCGTTT 600
Db 1630 GTGGCTCTGCTTTCTTCTTCATTTGACGAGTACTATATATACCAAGGGGATACACCGTTT 1689
Qy 601 GCTCCATA 608
Db 1690 GCTCCATA 1697

RESULT 4
AAA07075
ID AAA07075 standard; cDNA; 2372 BP.
AC AAA07075;
XX
XX
XX 03-JUL-2000 (first entry)
DT
XX
XX cDNA encoding human ataxin-2 binding protein (A2BP).
DE
XX
XX Ataxin-2 binding protein; A2BP; human; RNA-binding; cell polarisation;
KW neuronal plasticity; cellular degeneration signal transduction pathway;
KW selective RNA transport; spinocerebellar ataxia type-2;
KW hyperproliferative disorder; ss.
OS Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX CDS 987..1979
XX FT /*tag= a
XX FT /product= "Human A2BP"
XX
XX WO200012710-A1.
XX
XX 09-MAR-2000.
XX
XX 01-SEP-1999; 99WO-US20156.
XX
XX 01-SEP-1998; 98US-0145391.
XX
XX (CEDA-) CEDARS SINAI MEDICAL CENT.
XX
XX Pulst SM, Shibata H;
XX
XX MPI; 2000-237873/20.
XX
XX P-PSDB; AAY81462.
XX
XX Nucleic acids encoding an ataxin-2 binding protein useful for
XX inhibiting the expression of active proteins from the SCA2 gene for the
XX treatment of spinocerebellar ataxia type-2.
XX
XX Claim 6; Page 74-77; 82pp; English.
XX
XX This sequence represents cDNA encoding human ataxin-2 binding protein
XX (A2BP). Nucleotide sequences encoding human A2BP were originally
XX isolated in an adult brain cDNA library using the yeast two hybrid
XX method. The ligand of A2BP, ataxin-2, is a 40 kD protein of unknown
XX function that is encoded by the SCA2 gene located on chromosome 12. SCA2
XX has been linked to the autosomal dominant neurodegenerative disorder
XX spinocerebellar ataxia type-2. Individuals afflicted with the disease
XX exhibit CAG triplet expansion in the SCA2 gene, resulting in ataxin-2
XX containing a polyglutamine stretch of about 35-39 residues, whereas that
XX of normal individuals contains approximately 22 contiguous glutamine
XX residues. A2BP and ataxin-2 are components of a cellular degeneration
XX signal transduction pathway. The pathogenic expanded form of ataxin-2 has
XX a higher affinity for A2BP relative to normal ataxin-2; the presence of
XX the expanded form is likely to promote degeneration. A2BP and ataxin have
XX also been found to have a role in gene regulation. The binding of A2BP to
XX ataxin-2 plays an important role in controlling gene expression via the
```

```
CC targeting of transport of specific RNAs, selective RNA transport being
CC mediated via the RNA binding domains of A2BP. A2BP is expressed very
CC early in embryonic development. Both ataxin-2 and A2BP are able to bind
CC RNA, and are essential components of the RNA localisation network that
CC establishes cellular polarity in embryogenesis. In highly differentiated,
CC polarised cells such as neurons, A2BP and ataxin-2 have a similar
CC function and are required for neuronal plasticity. A2BP nucleic acids may
CC be used for the recombinant production of A2BP proteins or fragments
CC thereof according to standard methodologies. For example, an A2BP protein
CC with an ataxin-2 or RNA binding capability but no signal transduction
CC function can be used as a dominant negative inhibitor of the cellular
CC degeneration signal transduction pathway. A2BP proteins with a signal
CC transduction function can be used to treat hyperproliferative disorders
CC (e.g., cancer) via stimulation of the cellular degeneration pathway.
XX
XX
XX Sequence 2372 BP; 548 A; 726 C; 646 G; 452 T; 0 other;
SQ
Query Match 70.4%; Score 428.6; DB 21; Length 2372;
Best Local Similarity 82.0%; Pred. No. 8.2e-121;
Matches 542; Conservative 0; Mismatches 64; Indels 55; Gaps 2;
Qy 1 ATGACTAATATAAAGGCGGTGAACCCCTACACCAATGCTGGAATTAATCAGTTGTG 60
Db 1572 ATGACAAATATAAAGACCCCTCAACCTTATACAAATGCTGGAATTAATCAGTTGTG 1631
Qy 61 GCGCGGCTCTACAGCCCGGACCTTATGACGAGCAGGTGCTTTGCGCAGGCCAACCAG 120
Db 1632 GGTGCACTACAGTCCGGAATTTATGACGAGCAGGTGCTTTGCGCAGGCCAACCAG 1691
Qy 121 GAGGATCTTCCATATACAGTGGCCCGGCTTCACTTGTATATACTTTCGCAATGCTGCG 180
Db 1692 GAGGATCTTCCATATACAGTGGCCCGGCTTCACTTGTATATACTTTCGCAATGCTGCG 1751
Qy 181 TTTCATATCCGCGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
Db 1752 TTCCGCTATCCAGACGACCGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1811
Qy 241 GGTGCACCGGTATACACACCTTTCAGAGCTGGGCGGCCCGCCCAATCCGCGCTAT 300
Db 1812 GGTGCACCGGTATACACACCTTTCAGAGCTGGGCGGCCCGCCCGATCCGCGCTATC 1871
Qy 301 GCGGAGTATGTATACAGAGCCAGTATGAGCAATTAATTCATCAGGCTGCTACCT 360
Db 1872 GCGGAGTATGTATACAGAGCCAGTATGAGCAATTAATTCATCAGGCTGCTACCT 1929
Qy 361 GCATACCGGTATGAGCCAGCCAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 410
Db 1930 GCATACCGGTATGAGCCAGCCAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1989
Qy 411 -----CAGTTACGAGCAGTTT 427
Db 1990 TTGCTTCTGTTGACGAGATGAATTTCTGTATACACTGCTGACGTTACGAGCAGTTT 2049
Qy 428 ATGCTGCGGACCCCTACACCAACACACTTGTCCAGCCCGCCACCTACAGGCTTGTGCTCA 487
Db 2050 ATGCTGCGGACCCCTACACCAACACACTTGTCCAGCCCGCCACCTACAGGCTTGTGCTCA 2109
Qy 488 TGAATGCTTTGGCGCTTGAACGATGCAAGACTAGAGGACGATGATGAGGCTG 547
Db 2110 TGAATGCTTTGGCGCTTGAACGATGCAAGACTAGAGGACGATGATGAGGCTG 2169
Qy 548 TCGTTCTTTCTTCAATGAGGCTGATATATACCAAGGGGATACACCGTTTGTCTCAT 607
Db 2170 TCGTTCTTTCTTCAATGAGGCTGATATATACCAAGGGGATACACCGTTTGTCTCAT 2229
Qy 608 A 608
Db 2230 A 2230

RESULT 5
AAS93633
ID AAS93633 standard; cDNA; 481 BP.
```

Best Local Similarity 90.0%; Pred. No. 1,66-142;
Matches 547; Conservative 0; Mismatches 58; Indels 3; Gaps 1;
QY 1 ATGCTAATAAAGAGCGGTGAACCCCTACACCAATGGCTGGAATTAATCACTGTG 60
Db 806 ATGCAATATAAAGAGCGGTGAACCCCTATACAAATGGCTGGAATTAATCACTGTG 865
QY 61 GCGCGGTCTACAGCCCGGACTTCTATGACAGGACGGTCTGTGGCCAGGCCAAG 120
Db 866 GGTGAGCTACAGTCCGGAATTTCTATGACAGGACGGTCTGTGGCCAGGCCAAG 925
QY 121 GAGGATCTTCCATGTATACAGTGGCCCCAGTTCACTTGTATATACCTTTCGAATGCTTGGC 180
Db 926 GAGGATCTTCCATGTATACAGTGGCCCCAGTTCACTTGTATATACCTTTCGAATGCTTGGC 985
QY 181 TTTCCATATCCGGCGGCGGCACTGCTGAGTGCATACGAGGGGCTCACCTTTCGAGGCGCT 240
Db 986 TTCCCGTATCCAGGACGACCCGCGGCGGCTTACCGAGGGGCGCACTTTCGAGGCGCG 1045
QY 241 GGTGCAACCGTGTACACACCTTTCAGAGCTGGCGGCGGCGGCGGCGGCGGCGGCTAT 300
Db 1046 GGTGCAACCGTGTACACACCTTTCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCTAT 1105
QY 301 GCGGAGTATGTATCAAGAGCGGATGTATGCAATTAATGCTACAGGCTGTACCT 360
Db 1106 GCGGAGTATGTATCAAGAGCGGATGTATGCAATTAATGCTGTATGCT 1162
QY 361 GCATACCGCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTAT 420
Db 1163 GCATACCGCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTAT 1222
QY 421 CGAGTTATGTGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 480
Db 1223 CGAGTTATGTGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 1282
QY 481 GGTGCGATGAATGCTTTGGCGGCTTTCACCGATGCGAAGATGAGGAGCGGCTGATGAT 540
Db 1283 GGTGCGATGAATGCTTTGGCGGCTTTCACCGATGCGAAGATGAGGAGCGGCTGATGAT 1342
QY 541 GTGGGTCGTTCTTCTTCAATGAGGCTGATATATCAAGGGGGGATCAACGCTTTT 600
Db 1343 GTGGGTCGTTCTTCTTCAATGAGGCTGATATATCAAGGGGGGATCAACGCTTTT 1402
QY 601 GCTCCATA 608
Db 1403 GCTCCATA 1410
RESULT 3
AAK52245
ID AAK52245 standard; cDNA; 1800 BP.
XX
AC AAK52245;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide seq ID NO 790.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN MO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001MO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.

PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
XX
DR WPI: 2001-476283/51.
XX
PT P-PSDB: AAM79112.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 1; Page 2643-2645; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78333-AAM80302) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoietic regulatory
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 1800 BP; 432 A; 552 C; 481 G; 335 T; 0 other;
XX
Query Match 82.0%; Score 499.2; DB 22; Length 1800;
Best Local Similarity 90.0%; Pred. No. 1,76-142;
Matches 547; Conservative 0; Mismatches 58; Indels 3; Gaps 1;
QY 1 ATGCTAATAAAGAGCGGTGAACCCCTACACCAATGGCTGGAATTAATCACTGTG 60
Db 1093 ATGCAATATAAAGAGCGGTGAACCCCTATACAAATGGCTGGAATTAATCACTGTG 1152
QY 61 GCGCGGTCTACAGCCCGGACTTCTATGACAGGACGGTCTGTGGCCAGGCCAAG 120
Db 1153 GGTGAGCTACAGTCCGGAATTTCTATGACAGGACGGTCTGTGGCCAGGCCAAG 1212
QY 121 GAGGATCTTCCATGTATACAGTGGCCCCAGTTCACTTGTATATACCTTTCGAATGCTTGGC 180
Db 1213 GAGGATCTTCCATGTATACAGTGGCCCCAGTTCACTTGTATATACCTTTCGAATGCTTGGC 1272
QY 181 TTTCCATATCCGGCGGCGGCACTGCTGAGTGCATACGAGGGGCTCACCTTTCGAGGCGCT 240
Db 1273 TTCCCGTATCCAGGACGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1332
QY 241 GGTGCAACCGTGTACACACCTTTCAGAGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCTAT 300
Db 1333 GGTGCAACCGTGTACACACCTTTCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCTAT 1392
QY 301 GCGGAGTATGTATCAAGAGCGGATGTATGCAATTAATGCTACAGGCTGTACGCT 360
Db 1393 GCGGAGTATGTATCAAGAGCGGATGTATGCAATTAATGCTGTATGCT 1449
QY 361 GCATACCGCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTAT 420
Db 1450 GCATACCGCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTAT 1509
QY 421 CGAGTTATGTGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 480
Db 1510 CGAGTTATGTGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 1569

PA (SCIO-) SCIOS INC.
 XX Stanton LW, White RT;
 XX WPI: 2002-010779/01.
 DR P-PSDB; AAU70146.
 XX Novel secreted factor polypeptide useful for treating cardiac diseases
 PT such as arteriosclerosis, myocardial infarction, inflammatory diseases
 PT such as asthma, stroke, and rheumatoid arthritis and renal diseases -
 XX
 PS Claim 1; Fig 1; 189pp; English.
 XX The invention relates to rat secreted factor polypeptides and the
 CC polynucleotides encoding them. The sequences are useful for treating
 CC cardiac, renal or inflammatory diseases. These include cardiac diseases
 CC such as congestive heart failure, myocarditis, dilated congestive
 CC cardiomyopathy, angina pectoris, myocardial infarction, cardiac
 CC arrhythmia, pulmonary hypertension, arteriosclerosis, atherosclerosis and
 CC cardiac tumours, renal diseases such as glomerulonephritis, nephrotic
 CC syndrome, renal infarction, hereditary nephritis, polycystic kidney
 CC disease, chronic renal failure, renal vein thrombosis and medullary
 CC sponge kidney and inflammatory diseases such as asthma, rheumatoid
 CC arthritis, osteoarthritis, stroke, psoriasis, testenosis, graft versus
 CC host reaction, Crohn's disease, ulcerative colitis and Alzheimer's
 CC disease. Sequences AAS94693-AAS94745 represent cDNA clones, which encode
 CC the secreted factor polypeptides of the invention, and oligonucleotide
 CC probes and PCR primers.
 XX
 SQ Sequence 1340 BP; 366 A; 394 C; 317 G; 263 T; 0 other;
 Query Match 100.0%; Score 609; DB 24; Length 1340;
 Best Local Similarity 100.0%; Pred. No. 3.6e-176;
 Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 ATGACTAATAAAGAGCGCTGAACCCCTACACCAATGCTGGAATTAATCCAGTTGTG 60
 DB 535 ATGACTAATAAAGAGCGCTGAACCCCTACACCAATGCTGGAATTAATCCAGTTGTG 594
 OY 61 GGGGGGCTTACAGCCCGCACTTCTATGAGGAGCGTCTGTGTGCGAGGCAACGAG 120
 DB 595 GGGGGGCTTACAGCCCGCACTTCTATGAGGAGCGTCTGTGTGCGAGGCAACGAG 654
 OY 121 GAGGATCTTCATGATGACAGTGGCCCACTTCTATGATGATGATGATGATGATGAT 180
 DB 655 GAGGATCTTCATGATGACAGTGGCCCACTTCTATGATGATGATGATGATGATGAT 714
 OY 181 TTTCATATCCGGCCGCACTGCTGAGCTGCATACCGAGGGGCTACCTTGAGGCGCT 240
 DB 715 TTTCATATCCGGCCGCACTGCTGAGCTGCATACCGAGGGGCTACCTTGAGGCGCT 774
 OY 241 GGTCCGACCGTGAACAACCTTCAAGAGTGGCGGCCCGCCCAATCCCGGCTTAT 300
 DB 775 GGTCCGACCGTGAACAACCTTCAAGAGTGGCGGCCCGCCCAATCCCGGCTTAT 834
 OY 301 GCGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 DB 835 GCGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 894
 OY 361 GCATACCGTGAACCGCCGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
 DB 895 GCATACCGTGAACCGCCGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 954
 OY 421 CGAGTTATGCTGCGGACCCCTACCAACACACCTTGTCCAGCCCCCACTTACGCGCTT 480
 DB 955 CGAGTTATGCTGCGGACCCCTACCAACACACCTTGTCCAGCCCCCACTTACGCGCTT 1014
 OY 481 GGTGCGATGAATCTTTGGGCGCTTGAACCGATGCGAAGATGAGGATGATGATGAT 540
 DB 1015 GGTGCGATGAATCTTTGGGCGCTTGAACCGATGCGAAGATGAGGATGATGATGAT 1074
 OY 541 GTGGGTCTGCTTTCTTCTATGAGGCTAGTATATACCAAGGGGATACCAACGCTTTT 600

DB 1075 GTGGGTCTGCTTTCTTCTATGAGGCTAGTATATACCAAGGGGATACCAACGCTTTT 1134
 OY 601 GCTCCATAT 609
 DB 1135 GCTCCATAT 1143
 RESULT 2
 AAH13824
 ID AAH13824 standard; cDNA; 1513 BP.
 XX
 AC AAH13824;
 XX
 XX 26-JUN-2001 (first entry)
 DT
 DE Human cDNA sequence SEQ ID NO:10786.
 XX
 KM Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 OS Homo sapiens.
 PN EPI074617-A2.
 PD
 XX 07-FEB-2001.
 PF
 XX 28-JUL-2000; 2000EP-0116126.
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 XX Ota T, Teogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI: 2001-318749/34.
 PT
 PT primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 10786; 2537bp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 1513 BP; 354 A; 418 C; 395 G; 346 T; 0 other;
 Query Match 82.0%; Score 499.2; DB 22; Length 1513;

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 10:24:02 ; Search time 142.798 Seconds
(without alignments)
9604.255 Million cell updates/sec

Title: US-09-809-545a-1_COPY_535_1143
Perfect score: 609
Sequence: 1 atgactaataaaagccgt.....acaacgcttcgtccatcat 609

Scoring table:
IDENTITY NUC
Gapop 10-0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: N_GeneSeq_101002.*
2: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1980.DAT.*
3: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1981.DAT.*
4: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1982.DAT.*
5: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1983.DAT.*
6: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1984.DAT.*
7: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1985.DAT.*
8: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1986.DAT.*
9: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1987.DAT.*
10: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1988.DAT.*
11: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1989.DAT.*
12: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1990.DAT.*
13: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1991.DAT.*
14: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1992.DAT.*
15: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1993.DAT.*
16: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1994.DAT.*
17: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1995.DAT.*
18: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1996.DAT.*
19: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1997.DAT.*
20: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1998.DAT.*
21: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA1999.DAT.*
22: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA2000.DAT.*
23: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA2001A.DAT.*
24: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA2001B.DAT.*
25: /SID52/gcgdata/geneSeq/geneSeq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	609	100.0	1340	24	AAS94693
2	499.2	82.0	1513	22	AAH13824
3	499.2	82.0	1800	22	AAK52245
4	428.6	70.4	2372	21	AAAO7075
5	300.8	49.4	481	23	AAH93633
6	299.8	49.2	607	22	AAH9205
7	260	42.7	539	24	ABK34530
8	190	31.2	406	23	AAS93632
9	134.4	22.1	567	23	AAS66245

10	134.4	22.1	918	22	AA159249	Human polynucleoti
C 11	134.4	22.1	939	22	AA161034	Human polynucleoti
C 12	134.4	22.1	939	22	AA161035	Human polynucleoti
C 13	134.4	22.1	1011	22	AA159248	Human polynucleoti
C 14	133.6	21.9	327	22	AAK54001	Murine transcript
C 15	129	21.2	1164	22	AAK33229	Human polynucleoti
16	112	18.4	3189	24	ABK34602	Human cDNA for nov
17	107	17.6	2118	23	AAS93634	DNA encoding novel
18	73.4	12.1	1252	23	AAS90510	DNA encoding novel
19	63.2	10.4	1506	21	AA246827	Human RNA binding
20	55.2	9.1	60	24	ABN34672	Human spliced tran
21	47.8	7.8	578	20	AAK90506	DNA encoding novel
22	45	7.4	413	21	AAK04077	Human secreted pro
23	44.6	7.3	428	24	ABL68917	Kidney cancer rela
24	40.4	6.6	316	14	AAQ60335	Human brain Expres
C 25	35.8	5.9	2949	23	AAK67904	DNA encoding novel
C 26	35.6	5.8	577	20	AAK20782	Polynucleotide seq
C 27	35.6	5.8	4403765	22	AA199683	Mycobacterium seg
28	35.4	5.8	660	20	AAK37406	Mycobacterium tube
29	34.6	5.7	2564	23	ABL06127	Human secreted pro
30	34.6	5.7	5340	23	ABL06126	Drosophila melanog
31	34.6	5.7	12403	23	ABL07646	Drosophila melanog
32	34.2	5.6	15664	17	AAK09312	Mycobacteriophage
33	34.2	5.6	15664	18	AAK70492	Mycobacteriophage
34	34.2	5.6	15664	18	AAK6130	Mycobacteriophage
35	34.2	5.6	15664	18	AAK51224	NheI-G/SpeI fragme
C 36	34	5.6	3351	20	AAK109457	Drosophila melanog
C 37	34	5.6	3351	20	AAK109456	S. rochei strain B
C 38	34	5.6	3974	23	ABL09456	Drosophila melanog
C 39	33.4	5.5	538	19	AAK6456	M. tuberculosis im
C 40	33.4	5.5	538	19	AAK4457	Mycobacterium tube
C 41	33.4	5.5	538	20	AAK19347	M. tuberculosis an
C 42	33.4	5.5	538	20	AAK19135	M. tuberculosis re
C 43	33.4	5.5	2207	20	AAK77135	Human nerve mutat
C 44	33.4	5.5	10732	21	AAK10594	Gene encoding a su
45	33	5.4	1828	22	AAK1366	Peroxidase Swpa2 p

ALIGNMENTS

RESULT 1	
AAS94693	AAS94693 standard; cDNA; 1340 BP.
XX	
AC	AAS94693:
XX	
DT	12-MAR-2002 (first entry)
XX	
DE	Rat secreted factor DNA clone P0184_D11 #1.
XX	
KW	Rat: secreted factor polypeptide; cardiac disease; kidney;
KW	inflammatory disease; congestive heart failure; myocarditis; asthma; ss;
KW	dilated congestive cardiomyopathy; angina pectoris; cardiac arrhythmia;
KW	myocardial infarction; pulmonary hypertension; arteriosclerosis; stroke;
KW	atherosclerosis; cardiac tumor; glomerulonephritis; nephrotic syndrome;
KW	renal infarction; hereditary nephritis; polycystic kidney disease;
KW	chronic renal failure; renal vein thrombosis; medullary sponge kidney;
KW	rheumatoid arthritis; osteoarthritis; psoriasis; restenosis; PCR primer;
KW	glial virus host reaction; Crohn's disease; ulcerative colitis; probe;
KW	Alzheimer's disease; gene therapy.
XX	
OS	Rattus norvegicus.
XX	
PN	W0200174901-A2.
XX	
PD	11-OCT-2001.
XX	
PF	23-MAR-2001; 2001WO-US09555.
XX	
PR	31-MAR-2000; 2000US-193548P.
XX	
PR	14-MAR-2001; 2001US-0809545.
XX	

THIS PAGE BLANK (USPTO)

Job time : 1753.58 secs

```
repeat_region complement(99089, .99381)
repeat_region /rpt_family="Alu"
repeat_region 101648, .101930
repeat_region /rpt_family="Alu"
repeat_region 102790, .103056
repeat_region /rpt_family="Alu"
repeat_region 104010, .104277
repeat_region /rpt_family="Alu"
repeat_region complement(111121, .111263)
repeat_region /rpt_family="MER5"
repeat_region complement(112263, .112476)
repeat_region /rpt_family="MER30"
repeat_region 112956, .113234
repeat_region /rpt_family="Alu"
repeat_region 114385, .114657
repeat_region /rpt_family="Alu"
repeat_region complement(117031, .117194)
repeat_region /rpt_family="MER20"
repeat_region complement(118464, .118770)
repeat_region /rpt_family="Alu"
repeat_region 119271, .119549
repeat_region /rpt_family="Alu"
repeat_region complement(122831, .122940)
repeat_region /rpt_family="MER41"
repeat_region 123471, .123742
repeat_region /rpt_family="Alu"
repeat_region complement(129860, .129972)
repeat_region /rpt_family="Alu"
repeat_region 135235, .135310
repeat_region /rpt_family="MLT1"
repeat_region 135874, .136157
repeat_region /rpt_family="Alu"
repeat_region 139775, .140048
repeat_region /rpt_family="Alu"
repeat_region 141734, .141814
repeat_region /rpt_family="MIR"
repeat_region 143071, .143352
repeat_region /rpt_family="Alu"
repeat_region 148286, .148773
repeat_region /rpt_family="MER1"
repeat_region 150431, .150707
repeat_region /rpt_family="Alu"
repeat_region 155151, .155387
repeat_region /rpt_family="Alu"
repeat_region complement(156101, .156213)
repeat_region /rpt_family="MER20"
repeat_region complement(156356, .156920)
repeat_region /rpt_family="Alu"
repeat_region complement(159633, .159931)
repeat_region /rpt_family="Alu"
repeat_region complement(160378, .160658)
repeat_region /rpt_family="Alu"
repeat_region 161617, .162140
repeat_region /rpt_family="Alu"
repeat_region complement(163200, .163315)
repeat_region /rpt_family="MIR"

BASE COUNT 48114 a 34256 c 33738 g 48430 t
ORIGIN
```

```
Query Match 18.7% Score 113.6; DB 9; Length 164538;
Best Local Similarity 96.7% Pred. No. 1,1e-21;
Matches 116; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 489 GAATGCTTTGGCCCTTGACGATGCCAAGCTAGAGCCATGCTGATGATGCGGCT 548
Db 61350 GAATGCTTTGACCTTGACTGATGCCAAGCTAGAGCCATGCTGATGATGCGGCT 61291
Qy 549 CGTCTTCTTCATTCAGGCTAGATATACCAAGGGGATACAACGTTTGCTCCATA 608
Db 61290 CGTCTTCTTCATTCAGGCTAGATATACCAAGGGGATACAACGTTTGCTCCATA 61231
```

Search completed: March 15, 2003, 14:22:20

RESULT 15
AC005774/c 164538 bp DNA linear PRI 05-OCT-1998
LOCUS Homo sapiens chromosome 16, BAC clone 2603 (LANL), complete
DEFINITION sequence.
AC005774
VERSION AC005774.1 GI:3702264
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS Rieke, D.O.
TITLE 1 (bases 1 to 164538)
JOURNAL Large Scale Sequence Analysis and Annotation with the Sequence
REFERENCE Comparison Analysis (SCAN) System
AUTHORS Unpublished
JOURNAL 2 (bases 1 to 164538)
REFERENCE Rieke, D.O., Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E.,
AUTHORS Robinson, D., Jones, M., Buckingham, J., Chasteen, L., Thompson, S.,
Goodwin, L., Bryant, J., Tesmer, J., Meincke, L., Longmire, J.,
White, S., Ueng, S., Tatum, O., Campbell, C., Fawcett, J., Maltbie, M.
and Deaven, L.
TITLE Sequencing of Human Chromosome 16p13.3
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 164538)
AUTHORS Rieke, D.O., Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E.,
Robinson, D., Jones, M., Buckingham, J., Chasteen, L., Thompson, S.,
Goodwin, L., Bryant, J., Tesmer, J., Meincke, L., Longmire, J.,
White, S., Ueng, S., Tatum, O., Campbell, C., Fawcett, J., Maltbie, M.
and Deaven, L.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-1998) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
FEATURES
source
1..164538
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p13.3"
/clone="2603"
1378..1665
/rpt_family="Alu"
repeat_region
complement(2144..2320)
/rpt_family="MER21"
repeat_region
complement(2198..2416)
/rpt_family="MER39"
repeat_region
complement(2450..2725)
/rpt_family="Alu"
repeat_region
complement(3157..3393)
/rpt_family="Alu"
repeat_region
complement(3524..3792)
/rpt_family="Alu"
repeat_region
5731..6020
/rpt_family="Alu"
repeat_region
complement(7557..7727)
/rpt_family="MER4"
repeat_region
8282..8928
/rpt_family="LTR8"
repeat_region
10055..10310
/rpt_family="TBE1"
repeat_region
10472..10618
/rpt_family="L1"
repeat_region
10639..10939
/rpt_family="Alu"
repeat_region
complement(11109..11357)
/rpt_family="Alu"
repeat_region
12050..12312
/rpt_family="Alu"
repeat_region
complement(14642..14755)

repeat_region
/rpt_family="L1"
17429..18370
/rpt_family="MER21"
repeat_region
complement(19470..19819)
/rpt_family="MER21"
repeat_region
complement(20352..20627)
/rpt_family="Alu"
repeat_region
23808..24058
/rpt_family="Alu"
repeat_region
27654..28217
/rpt_family="L1"
repeat_region
complement(32764..32956)
/rpt_family="L1"
repeat_region
complement(33392..33503)
/rpt_family="Alu"
repeat_region
34586..34870
/rpt_family="Alu"
repeat_region
34936..35489
/rpt_family="MLT2B2"
repeat_region
complement(36940..37172)
/rpt_family="Alu"
repeat_region
37173..37437
/rpt_family="TBE1"
repeat_region
41428..41745
/rpt_family="TBE1"
repeat_region
complement(42057..43927)
/rpt_family="L1"
repeat_region
43924..44555
/rpt_family="L1"
repeat_region
45128..45667
/rpt_family="Alu"
repeat_region
45467..45704
/rpt_family="Alu"
repeat_region
46192..46578
/rpt_family="MER25"
repeat_region
complement(46779..47141)
/rpt_family="TBE1"
repeat_region
complement(48265..48550)
/rpt_family="Alu"
repeat_region
49630..50221
/rpt_family="MER41"
repeat_region
50781..51096
/rpt_family="Alu"
repeat_region
complement(51219..51439)
/rpt_family="Alu"
repeat_region
51462..51708
/rpt_family="MER33"
repeat_region
complement(52042..52321)
/rpt_family="Alu"
repeat_region
54001..54092
/rpt_family="Alu"
repeat_region
complement(55760..56014)
/rpt_family="Alu"
repeat_region
56616..56899
/rpt_family="Alu"
repeat_region
complement(66463..66747)
/rpt_family="Alu"
repeat_region
complement(66908..67180)
/rpt_family="Alu"
repeat_region
complement(70389..70704)
/rpt_family="MER1"
repeat_region
complement(72166..72421)
/rpt_family="MER33"
repeat_region
72501..72769
/rpt_family="Alu"
repeat_region
72843..72956
/rpt_family="MER5"
repeat_region
complement(76303..76564)
/rpt_family="Alu"
repeat_region
97534..97700
/rpt_family="MER3"
repeat_region
complement(98469..98526)
/rpt_family="Alu"

SOURCE	
ORGANISM	Homo sapiens.
REFERENCE AUTHORS TITLE	Homologous to fox-1 in Caenorhabditis elegans Chen, W., Chu, Z.-L., Blough, R.I., Liu, L., Hoppes, B. and Winkelmann, J.C. Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 2971) The Exon-Intron Organization of the Human HRNBP1 Gene Unpublished 2 (bases 1 to 2971)
JOURNAL REFERENCE AUTHORS TITLE	Submitted (25-NOV-1998) Internal Medicine/Hematology-Oncology, University of Cincinnati College of Medicine, 231 Bethesda Ave. ML 0508, Cincinnati, OH 45267-0508, USA Location/Qualifiers 1. 2971
FEATURES source	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="16" /map="16p13.3" order(AFI09107..1:372..948,AFI09108..1:1..470, AFI09109..1:1..629,AFI09110..1:1..219,AFI09111..1:1..219, AFI09112..1:1..253,AFI09113..1:1..347,AFI09114..1:1..374, AFI09115..1:1..251,AFI09116..1:1..249,AFI09117..1:1..151, AFI09118..1:1..221,AFI09119..1:1..234,AFI09120..1:1..2033) /gene="HRNBP1" join(AFI09107..1:372..711,AFI09108..1:173..415, AFI09109..1:166..309,AFI09110..1:94..147,AFI09111..1:75..167, AFI09112..1:50..110,AFI09113..1:200..253,AFI09114..1:22..102, AFI09115..1:67..199,AFI09116..1:108..150,AFI09117..1:35..99, AFI09118..1:124..176,AFI09119..1:62..137,AFI09120..1:35..99, AFI09121..1:159..2033) /product="hexaribonucleotide binding protein 1 isoform beta"
gene	
MRNA	
MRNA	
CDs	

KGLIDVEILFNERGSGKGFVTEPENSADYADRAAREKHLGTVDEGRKLEIVNNATARVMVNN
 KTYVNPPTNNMKNLNPVVGAYVSEPFYVNTGATVLLCOANOEGSSMSAPESLVTYSAMPF
 FYPAIDTAALVRCGHLERGRGRIVYNTFRAPAAPPEIYATGCVYQDEYVGNKLLQGGT
 AARVYAOPTPATAAVYSDSYGRVYAADPYNHAAAPATYGVGMNNAFAPLLDIAKTRSH
 ADDVGLVSLLSQASIVYGGYNRRAPY"
 join(AFI09107.1:625, .711,AFI09108.1:1173, .415,
 AFI09109.1:166, .309,AFI09110.1:94, .147,AFI09111.1:75, .1
 AFI09112.1:50, .110,AFI09113.1:200, .253,AFI09114.1:22, .1
 AFI09115.1:67, .199,AFI09116.1:108, .150,AFI09117.1:35, .9
 AFI09119.1:62, .137,237, .281)
 /gene="HNRBP1"
 /note="similar to Caenorhabditis elegans fox-1;
 alternatively spliced"
 /codon_start=1
 /product="hexaribonucleotide binding protein 1 isoform
 alpha"
 /protein_id="AAU83408.1"
 /db_xref="GI:19032385"
 /translation="MLASQGVLLHPYGVPMIVPAAPYLPCLIGNOGNAAPPTMOP
 YASAOFPONGIYPAETAPHPAPAPETQOTVPEHTLILYPAQCHSOSPADISA
 QYVSGATQTDADAPDQOPOTOPSENTENKSKPKLHVSNIPREFPDRLRMFGQF
 GKLIDVEILFNERGSGKGFVTEPENSADYADRAAREKHLGTVDEGRKLEIVNNATARVMN
 KTYVNPPTNNMKNLNPVVGAYVSEPFYVNTGATVLLCOANOEGSSMSAPESLVTYSAMPF
 FYPAIDTAALVRCGHLERGRGRIVYNTFRAPAAPPEIYATGCVYQDEYVGNKLLQGGY
 AARVYAOPTPATAAVYSDSYGRVYAADPYNHAAAPATYGVGMNNAFAPLLDIAKTRSH
 ADDVGLVSLLSQASIVYGGYNRRAPY"
 join(AFI09107.1:625, .711,AFI09108.1:1173, .415,
 AFI09109.1:166, .309,AFI09110.1:94, .147,AFI09111.1:75, .1
 AFI09112.1:50, .110,AFI09113.1:200, .253,AFI09114.1:22, .1
 AFI09115.1:67, .199,AFI09116.1:108, .150,AFI09117.1:35, .9
 AFI09119.1:62, .137,237, .281)
 /gene="HNRBP1"
 /note="similar to Caenorhabditis elegans fox-1;
 alternatively spliced"
 /codon_start=1
 /product="hexaribonucleotide binding protein 1 isoform
 beta"
 /protein_id="AAU83407.1"
 /db_xref="GI:19032384"
 /translation="MLASQGVLLHPYGVPMIVPAAPYLPCLIGNOGNAAPPTMOP
 YASAOFPONGIYPAETAPHPAPAPETQOTVPEHTLILYPAQCHSOSPADISA
 QYVSGATQTDADAPDQOPOTOPSENTENKSKPKLHVSNIPREFPDRLRMFGQF
 GKLIDVEILFNERGSGKGFVTEPENSADYADRAAREKHLGTVDEGRKLEIVNNATARVMN
 KTYVNPPTNNMKNLNPVVGAYVSEPFYVNTGATVLLCOANOEGSSMSAPESLVTYSAMPF
 FYPAIDTAALVRCGHLERGRGRIVYNTFRAPAAPPEIYATGCVYQDEYVGNKLLQGGY
 AARVYAOPTPATAAVYSDSYGRVYAADPYNHAAAPATYGVGMNNAFAPLLDIAKTRSH
 ADDVGLVSLLSQASIVYGGYNRRAPY"
 join(AFI09107.1:625, .711,AFI09108.1:1173, .415,
 AFI09109.1:166, .309,AFI09110.1:94, .147,AFI09111.1:75, .1
 AFI09112.1:50, .110,AFI09113.1:200, .253,AFI09114.1:22, .1
 AFI09115.1:67, .199,AFI09116.1:108, .150,AFI09117.1:35, .9
 AFI09119.1:62, .137,237, .281)
 /gene="HNRBP1"
 /note="similar to Caenorhabditis elegans fox-1;
 alternatively spliced"
 /codon_start=1
 /product="hexaribonucleotide binding protein 1 isoform
 beta"
 /protein_id="AAU83407.1"
 /db_xref="GI:19032384"
 /translation="MLASQGVLLHPYGVPMIVPAAPYLPCLIGNOGNAAPPTMOP
 YASAOFPONGIYPAETAPHPAPAPETQOTVPEHTLILYPAQCHSOSPADISA
 QYVSGATQTDADAPDQOPOTOPSENTENKSKPKLHVSNIPREFPDRLRMFGQF
 GKLIDVEILFNERGSGKGFVTEPENSADYADRAAREKHLGTVDEGRKLEIVNNATARVMN
 KTYVNPPTNNMKNLNPVVGAYVSEPFYVNTGATVLLCOANOEGSSMSAPESLVTYSAMPF
 FYPAIDTAALVRCGHLERGRGRIVYNTFRAPAAPPEIYATGCVYQDEYVGNKLLQGGY
 AARVYAOPTPATAAVYSDSYGRVYAADPYNHAAAPATYGVGMNNAFAPLLDIAKTRSH
 ADDVGLVSLLSQASIVYGGYNRRAPY"
 join(AFI09107.1:625, .711,AFI09108.1:1173, .415,
 AFI09109.1:166, .309,AFI09110.1:94, .147,AFI09111.1:75, .1
 AFI09112.1:50, .110,AFI09113.1:200, .253,AFI09114.1:22, .1
 AFI09115.1:67, .199,AFI09116.1:108, .150,AFI09117.1:35, .9
 AFI09119.1:62, .137,237, .281)
 /gene="HNRBP1"
 /note="similar to Caenorhabditis elegans fox-1;
 alternatively spliced"
 /codon_start=1
 /product="hexaribonucleotide binding protein 1 isoform
 beta"
 /protein_id="AAU83407.1"
 /db_xref="GI:19032384"
 /translation="MLASQGVLLHPYGVPMIVPAAPYLPCLIGNOGNAAPPTMOP
 YASAOFPONGIYPAETAPHPAPAPETQOTVPEHTLILYPAQCHSOSPADISA
 QYVSGATQTDADAPDQOPOTOPSENTENKSKPKLHVSNIPREFPDRLRMFGQF
 GKLIDVEILFNERGSGKGFVTEPENSADYADRAAREKHLGTVDEGRKLEIVNNATARVMN
 KTYVNPPTNNMKNLNPVVGAYVSEPFYVNTGATVLLCOANOEGSSMSAPESLVTYSAMPF
 FYPAIDTAALVRCGHLERGRGRIVYNTFRAPAAPPEIYATGCVYQDEYVGNKLLQGGY
 AARVYAOPTPATAAVYSDSYGRVYAADPYNHAAAPATYGVGMNNAFAPLLDIAKTRSH
 ADDVGLVSLLSQASIVYGGYNRRAPY"
 join(AFI09107.1:625, .711,AFI09108.1:1173, .415,
 AFI09109.1:166, .309,AFI09110.1:94, .147,AFI09111.1:75, .1
 AFI09112.1:50, .110,AFI09113.1:200, .253,AFI09114.1:22, .1
 AFI09115.1:67, .199,AFI09116.1:108, .150,AFI09117.1:35, .9
 AFI09119.1:62, .137,237, .281)
 /gene="HNRBP1"
 /note="similar to Caenorhabditis elegans fox-1;
 alternatively spliced"
 /codon_start=1
 /product="hexaribonucleotide binding protein 1 isoform
 beta"
 /protein_id="AAU83407.1"
 /db_xref="GI:19032384"
 /translation="MLASQGVLLHPYGVPMIVPAAPYLPCLIGNOGNAAPPTMOP
 YASAOFPONGIYPAETAPHPAPAPETQOTVPEHTLILYPAQCHSOSPADISA
 QYVSGATQTDADAPDQOPOTOPSENTENKSKPKLHVSNIPREFPDRLRMFGQF
 GKLIDVEILFNERGSGKGFVTEPENSADYADRAAREKHLGTVDEGRKLEIVNNATARVMN
 KTYVNPPTNNMKNLNPVVGAYVSEPFYVNTGATVLLCOANOEGSSMSAPESLVTYSAMPF
 FYPAIDTAALVRCGHLERGRGRIVYNTFRAPAAPPEIYATGCVYQDEYVGNKLLQGGY
 AARVYAOPTPATAAVYSDSYGRVYAADPYNHAAAPATYGVGMNNAFAPLLDIAKTRSH
 ADDVGLVSLLSQASIVYGGYNRRAPY"
 join(AFI09107.1:625, .711,AFI09108.1:1173, .415,
 AFI09109.1:166, .309,AFI09110.1:94, .147,AFI09111.1:75, .1
 AFI09112.1:50, .110,AFI09113.1:200, .253,AFI09114.1:22, .1
 AFI09115.1:67, .199,AFI09116.1:108, .150,AFI09117.1:35, .9
 AFI09119.1:62, .137,237, .281)
 /gene="HNRBP1"
 /note="similar to Caenorhabditis elegans fox-1;
 alternatively spliced"
 /codon_start=1
 /product="hexaribonucleotide binding protein 1 isoform
 beta"
 /protein_id="AAU83407.1"
 /db_xref="GI:19032384"
 /translation="MLASQGVLLHPYGVPMIVPAAPYLPCLIGNOGNAAPPTMOP
 YASAOFPONGIYPAETAPHPAPAPETQOTVPEHTLILYPAQCHSOSPADISA
 QYVSGATQTDADAPDQOPOTOPSENTENKSKPKLHVSNIPREFPDRLRMFGQF
 GKLIDVEILFNERGSGKGFVTEPENSADYADRAAREKHLGTVDEGRKLEIVNNATARVMN
 KTYVNPPTNNMKNLNPVVGAYVSEPFYVNTGATVLLCOANOEGSSMSAPESLVTYSAMPF
 FYPAIDTAALVRCGHLERGRGRIVYNTFRAPAAPPEIYATGCVYQDEYVGNKLLQGGY
 AARVYAOPTPATAAVYSDSYGRVYAADPYNHAAAPATYGVGMNNAFAPLLDIAKTRSH
 ADDVGLVSLLSQASIVYGGYNRRAPY"
 join(AFI09107.1:625, .711,AFI09108.1:1173, .415,
 AFI09109.1:166, .309,AFI09110.1:94, .147,AFI09111.1:75, .1
 AFI09112.1:50, .110,AFI09

Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hughes, M., Holloway, C., Hollins, B., Homai, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratoch, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louisgeed, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M. P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, R., Morris, S., Moser, N., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogun, M., Okunolu, G., Oregunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Siscon, I., Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Tellod, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinsom, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

TITLE
Journal Submission
Unpublished
2 (bases 1 to 115860)

JOURNAL
Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 115860)

REFERENCE
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 18, 2002 this sequence version replaced gi:20514522.

AUTHORS
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

JOURNAL
Project Information
Center project name: GMLC
Center clone name: CH230-34C2
Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 99693 bases at least Q40
Consensus quality: 101936 bases at least Q30
Consensus quality: 102743 bases at least Q20

COMMENT
NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 21 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1347: contig of 1347 bp in length
1348 1447: gap of unknown length
1448 2834: contig of 1387 bp in length

2835 2934: gap of unknown length
2935 4253: contig of 1319 bp in length
4254 4353: gap of unknown length
4354 5477: contig of 1124 bp in length
5478 5577: gap of unknown length
5578 7484: contig of 1907 bp in length
7485 7585: gap of unknown length
7586 9406: contig of 1822 bp in length
9407 9506: gap of unknown length
9507 11712: contig of 2206 bp in length
11713 11812: gap of unknown length
11813 13462: contig of 1650 bp in length
13463 13562: gap of unknown length
13563 17222: contig of 3660 bp in length
17223 17322: gap of unknown length
17323 20617: contig of 3295 bp in length
20618 20717: gap of unknown length
20718 24825: contig of 4108 bp in length
24826 24925: gap of unknown length
24926 28811: contig of 3886 bp in length
28812 33686: contig of 4775 bp in length
33687 33786: gap of unknown length
33787 40217: contig of 6431 bp in length
40218 40317: gap of unknown length
40318 47912: contig of 7595 bp in length
47913 48012: gap of unknown length
48012 55482: contig of 8470 bp in length
55483 56483: gap of unknown length
56484 66204: contig of 9622 bp in length
66205 66305: gap of unknown length
66306 74546: contig of 8242 bp in length
74547 74647: gap of unknown length
74648 86522: contig of 11876 bp in length
86523 100049: gap of unknown length
100050 100149: contig of 13427 bp in length
100150 115860: gap of unknown length
115861 15711: contig of 15711 bp in length.

FEATURES
source 1..115860
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-34C2"

BASE COUNT 33039 a 22417 c 22786 g 30574 t 7044 others

ORIGIN

Query Match 19.6%; Score 119.4; DB 2; Length 115860;
Best Local Similarity 99.2%; Pred. No. 2.2e-23;
Matches 120; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 489 GATGCTTTGGCCCTTGACCGATGCCAAGACTAGAGCCATGCTGATGATGGGTCT 548
DB 112555 GATGCTTTGGCCCTTGACCGATGCCAAGACTAGAGCCATGCTGATGATGGGTCT 112496

QY 549 CATTCTTTCTTATGACGCTAGATATATCCAGAGGGGATCAACCGTTTGTCCATA 608
DB 112495 CATTCTTTCTTATGACGCTAGATATATCCAGAGGGGATCAACCGTTTGTCCATA 112436

QY 609 T 609

DB 112435 T 112435

RESULT 14
HSFOX14
LOCUS HSFOX14 2971 bp DNA linear PRI 01-MAR-2002
DEFINITION Homo sapiens hexarionucleotide binding protein 1 alpha, beta, and gamma isoforms (HRNP1) gene, alternatively spliced, exon 14 and partial cds and complete cds.
ACCESSION AF109120
VERSION AF109120.1 GI:19032383
KEYWORDS
SEGMENT 14 of 14

QY	61	GGCGGGGCTACAGCGCCCGACCTTCTATGAGGACGGTCTGTGGCGCAGGCAACAG	120
Db	1632	GGTCAGCTACAGTCCGCAATTCTATGAGGACGGTCTGTGGCGCAGGCAACAG	1691
QY	121	GAGGGATCTTCATGTACAGTGGCCCCAGTTCACTTTATATATCTTCTGCAATGCTGGC	180
Db	1692	GAGGGATCTTCATGTACAGTGGCCCCAGTTCACTTTATATATCTTCTGCAATGCGAGGC	1751
QY	181	TTTCCATATCCGGCGGGCCACTGTGAGTGAATCCGAGGGGGCTCACTTGTGAGGGCGT	240
Db	1752	TTCCGATATCCAGCAGCCACCGCCGGCGGCTCTACCGAGGGGCGCACTTGCGAGGGCGC	1811
QY	241	GGTCGACCGGTGTACAACACTTTCAGAGCTGGGGCCCCCAACCCCAATCCCGGCTAT	300
Db	1812	GGTCGACCGGTGTACAACACTTTCAGAGGCGGGGGCCCCCGCCATCCCGGCTAC	1871
QY	301	GGCGGAGTATGTATACAGAGCCAGTGTATGCGAATAATTCCTACAGGGTGTATCGCT	360
Db	1872	GGCGGTGTGTATTAC--CAGATGGAATTTATGTGCGAGACATTTATGTGTATGCT	1929
QY	361	GCATACCGGTACGGCCAGGCCACCCCTGCGACTGCTGCTCAACAGTA-----	410
Db	1930	GCATACCGGTACGGCCAGGCTTACCCCTGCCACTGCGCTTACAGTACGAATAACG	1989
QY	411	-----CAGTTACGAGCAGTT	427
Db	1990	TTTCGTTCTGTGACAGAGATGAATTTCTTGAACACTCTGCAGTTACGAGCAGGTTT	2049
QY	428	ATGCTGCGGACCCCTACCAACCAACACTTGTCTCCAGCCCCCACTAAGCGCTTTGGTCCA	487
Db	2050	ATGCTGCGGACCCCTACCAACCAACACTTGTCTCCAGCCCCCACTAAGCGCTTTGGTCCA	2109
QY	488	TGAATGCTTTTGGCGCCTTGACCGATGCGCAACTAGAGCCATGCTGATGATGTGGGTC	547
Db	2110	TGAATGCTTTTGGCACTTGACTGATGATGCCAAGACTAGAGCCATGCTGATGATGTGGGTC	2169
QY	548	TGCTTCTTTCTCATTTGACAGGCTAGTATATACCAAGGGGATACAACGTTTGTCTCAT	607
Db	2170	TGCTTCTTTCTCATTTGACAGGCTAGTATATACCAAGGGGATACAACGTTTGTCTCAT	2229
QY	608	A 608	
Db	2230	A 2230	

RESULT	12
LOCUS	HSM803066
DEFINITION	Homo sapiens mRNA; CDNA DKFZP547L059 (from clone DKFZP547L059).
ACCESSION	AL713700
VERSION	AL713700.1 GI:19584415
KEYWORDS	.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 1623) Bloecher,H., Boecher,M., Brandt,P., Mewes,H.W., Weil,B. and Wiemann,S. Direct Submission Submitted (11-MAR-2002) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by GBP (National Research Centre for Biotechnology Ltd., Brunnschweig/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZP547L059) is available at the RZPD in Berlin. Please contact the RZPD: Resourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, Germany; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://mips.gsf.de/proj/cdna/.
TITLE	JOURNAL
COMMENT	

FEATURES	source	Location/Qualifiers
		1..1623
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/map="16p13.3"
		/clone="DKFZp547L059"
		/cissue type="brain"
		/clone_id="547 (synonym: hfbri). Vector pSport1; host DH10B; sites NotI + SalI"
		/dev_stage="fetal"
gene		1..1623
		/gene="DKFZp547L059"
		836..>1623
CDS		/gene="DKFZp547L059"
		/note="FLJ10165 fig"
		/codon_start=1
		/product="hypothetical protein"
		/protein_id="CAD28499.1"
		/db_xref="GI:19584416"
		/translation="MNCERLPLRGNOEAAAPDTMAQPYASOAPFAPONGIPAEYTA RHPHAPRYTGCTVPEHTLNLYPAGTHSOSPADTSAGTSGATOTDDAIPADG QTQSEATENKSKPKRLHVNIPRPDPDRLQMGFGPKLIDVEIIFNEGSGRGG VTFENSDADRARLKLHGTVEGRKLEVNATATVMNKKTVNPTYGMKLNIPVGA VSPFVAGTLLCANDEGSSMVASPFLVTSAMPGFYPAATA"
BASE COUNT	356 a 523 c 487 g 257 t	
ORIGIN		
Query Match	28.3%	Score 172.6; DB 9; Length 1623;
Best Local Similarity	90.6%	Pred. No. 5.1e-39;
Matches 184; Conservative	0; Mismatches 19; Indels 0; Gaps 0;	
QY	1	ATGACTATAAAAAAGGCGGTGACCCCTCAACCAATGAGTGGCTGGAATTAATCAAGTTGG 60
Db	1421	ATGCAAAATATAAAGACCGCTCAACCTTATATCAAAATGCTGGAATTAATCAAGTTGTG 1480
QY	61	GGCGCGTCTACAGCGCCGCACTTCTATGCAAGCAAGGTGCTGTTGTGCGAGGCCACAG 120
Db	1481	GGTGACGCTACAGTCCGGAATTCTATGCAAGCAAGGTCTGTTGTGCGAGGCCACAG 1540
QY	121	GAGGATCTTCCATGATACAGTGGGCCGATTCACCTTGATATATCTTCGCAATGCTTGGC 180
Db	1541	GAGGATCTTCCATGATACAGTGGGCCGATTCACCTTGATATATCTTCGCAATGCGAGC 1600
QY	181	TTTCATATCGGCGGCGCACTGC 203
Db	1601	TTCCGATATCGAGCGCCACCGC 1623

RESULT 13					
AC120661/c					
LOCUS	AC120661	115860 bp	DNA	linear	HTG 23-JUN-2002
DEFINITION	Rattus norvegicus clone CH230-34C2, *** SEQUENCING IN PROGRESS ***.				
	21 unordered pieces.				
ACCESSION	AC120661				
KEYWORDS	AC120661.2	GI:21902877			
SOURCE	HTG; HTGS_PHASE1.				
ORGANISM	Rattus norvegicus.				
REFERENCE	Rattus norvegicus.				
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
	1 (bases 1 to 115860)				
	Muzny, D.M., Adams, C., Adio-Oduola, B., All- osman, F.R., Allen, C., Alsbrook, S.L., Amaratunga, H.C., Are, J.R., Ayala, M., Banks, T., Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bonck, J., Bowie, S., Briefe, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dahorne, S.R., David, R., Davis, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,				

QY	1	ATGACATAAAAAAGGCGGTGAACCCCTACACCAATGGCTGGAAATTAATCCAGTTGG	60
Db	899	ATGACATAATAAAAAAGCCTCAACCCCTTATACAAATAGCTGGAAATTAATCCAGTTGG	958
QY	61	GGCGCGCTTACAGCGCCCGCATCTTATATGACAGACACGCTGCTTGTGCCAGGCCAACCG	120
Db	959	GGTGCAGCTCTACATGCTCCGAAATTCTATATGACAGCACGCTCTGTGTGGCCAGGCCAACCG	1018
QY	121	GAGGGATCTTCCATGTACACAGTGGGCCCACTTACCTTGTATATACCTTCTGCAATGCTTGGC	180
Db	1019	GAGGGATCTTCCATGTACAGTGGGCCCACTTACCTTGTATATACCTTCTGCAATGCTTGGC	1078
QY	181	TTTCCATATCCGGCGGCCCATCTGTGACGTGCATACCGAGGGAGTCCACCTTGAGAGCCGT	240
Db	1079	TTTCCGCTATCCACGACGCCACCGCCGGCGGCTTACCGAGGGAGTCCACCTTGAGAGCCGT	1138
QY	241	GCTCGCACCGTGTACAAACACTTTCAGAGCTGCGCGGCCCCCAACCCCAATCCCGGCTAT	300
Db	1139	GGTCCGACCGTGTACAAACACTTTCAGAGGGCGCGGGGCCCGCCCGCCGATCCCGGCTATC	1198
QY	301	GGCGGAGTATGTATATCAGAGCCAGTGTATGGCAATAAATGCTACAGGGTGTATACCTT	360
Db	1199	GGCGGAGTATGTATATCAGAGCCAGTGTATGGCAATAAATGCTACAGGGTGTATATGCT	1258
QY	361	GCATACCGGTACGCCACAGCCGCCCTGTGACGTGCGTACGCTACAGTGA-----	410
Db	1259	GCATACCGGTACGCCACAGCTTACCTGCTGACGTGCGCTACGCTACAGTGAACAAATCAG	1318
QY	411	-----CAGTTACGACGAGTTT	427
Db	1319	TTTCGTCTTGTGACGACAGATGAATTTCTTGTAAACACTGTGACGTTATCGACGAGTTT	1378
QY	428	ATGCTGCCGACCCCTTACCAACAACAATTGCTTCACAGCCCCCACTTACGAGCGTTGTGCCA	487
Db	1379	ATGCTGCCGACCCCTTACCAACAACAAGCTTGTCTCCAGCCCCCACTTACGAGCGTTGTGCCA	1438
QY	488	TGAATGCTTTTGGCGGCTTGACCGAATGCCAAGACTAGAGAGCATGTATGATGATGGGCTC	547
Db	1439	TGAATGCTTTTGGCACCTTTGACTGTAGATGCCAAGACTAGAGAGCATGTATGATGATGGGCTC	1498
QY	548	TGCTTCTTCTTTCATTTGACAGCGCTAGTATATACCAAGGGAGATACAAACGTTTGTCTCAT	607
Db	1499	TGCTTCTTCTTTCATTTGACAGCGCTAGTATATACCAAGGGAGATACAAACGTTTGTCTCAT	1558
QY	608	A 608	
Db	1559	A 1559	
RESULT 9			
AB060859			
LOCUS			
DEFINITION	AB060859	1475 bp mRNA linear PRI 13-JUN-2001	
ACCESSION	AB060859	Macaca fascicularis brain cDNA clone:Qcra-11594, full insert sequence.	
VERSION	AB060859.1	GI:13874510	
KEYWORDS		o1igo capping; f1s (full insert sequence).	
SOURCE		Macaca fascicularis adult male temporal lobe r1ght cDNA co mRNA, clone 11b;macaque brain cDNA library Qcra clone:Qcra-11594.	
ORGANISM		Macaca fascicularis	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae; Cercopitheciinae; Macaca.	
AUTHORS		1 (bases)	
TITLE		Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., H1rai, M., Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K.	
JOURNAL		Isolation of full-length cDNA clones from macaque brain cDNA libraries	
REFERENCE		2 (bases 1 to 1475)	
AUTHORS		Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.	
TITLE		Direct Submission	
JOURNAL		Submitted (27-Apr-2001) Katsuyuki Hashimoto, National Institute of	

Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
 (E-mail: khaethi@nih.go.jp, URL: http://www.nih.go.jp/yoken/genecbank/, Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)
 Lab host: TOP10
 Vector: pME18S-FLJ3 (Acc.No. AB009864)
 R. Site1: DraIII (CACTGTGTG)
 R. Site2: DraIII (CACCATGG)
 Description: 1st strand cDNA was primed with an oligo (dG) primer [ATGTCGCGCTTTTGTTTTGTTTT]; double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18S-FLJ3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by oligo-capping method (Sugano et al., Institute of Medical Science, University of Tokyo).
 Custom primer used for sequencing
 (5' end primer [CTTTCGCTCTTAAGACGCG] ;
 3' end primer [CGACCTGACGCTCAGCAGCA]).
 Location/Qualifiers
 1. 1475
 /organism="Macaca fascicularis"
 /db_xref="taxon:9541"
 /clone="Oetra-11594"
 /sex="male"
 /tissue_type="temporal lobe right"
 /clone_lib="macaque brain cDNA library Oetra"
 /dev_stage="adult"
 90. 1220
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="BAB66877.1"
 /db_xref="GI:13874511"
 /translation="MEKSGSRMWQGNQAEAAAPDTMAQPYASNQEPFNGIPAEYTT
 APHPHPAEYTGQTVPEHLNLNLPAPQOTSEDSPATNQTQVSGTATQTDAAPTDGO
 OPOPTPEENTENKSOPRLVHLSNIPERFROPDIPROMGQSGKILVDIIFNERSKGF
 GFPTFENSADADBARERKLTGTVVEGRKIEVNNAITARMTNKTIVNPLVGNKLPVVG
 AVSPERFAGTVLTCOMNOGSSWYSPSLVYTSAMGPPPYPAATAAAYRGALHRLG
 RGVTVNTPFAAAPPRPIPAVAGGVVDDGPGVADIVGVAAYRVAQOPTPATNAAVSDR
 NQPFVAADIEICNTSAVTDPEMLPPTTHLLOPPTALVP"
 BASE COUNT 389 a 414 c 363 g 309 t
 ORIGIN
 Query Match 71.1%; Score 433; DB 9; Length 1475;
 Best Local Similarity 82.5%; Pred. No. 1,1e-114;
 Matches 545; Conservative 0; Mismatches 60; Indels 56; Gaps 2;
 QY 1 ATGACTAATAAAGAGCCGTGAACCCCTTACACCAATGCGTGGAAATTAATTCAGTTGTG 60
 Db 681 ATGACAAATAAAGAGACCGGTCAACCCCTTATACAAATGCGTGGAAATTAATTCAGTTGTG 740
 QY 61 GGGCGGCTCTACAGCCCGCGAATTCTATGACAGGACGGCGTGTGTGCGACGCAACAG 120
 Db 741 GGTGGGGCTACAGTCGGGAATCTATGACAGGACGGCGTGTGTGCGACGCAACAG 800
 QY 121 GAGGATTTTCATGATACAGTGGGCCCGCCAGTTCACTTGATATTACTTCGCAATGCTGGC 180
 Db 801 GAGGATTTTCATGATACAGTGGGCCCGCCAGTTCACTTGATATTACTTCGCAATGCAAGC 860
 QY 181 TTTCCATATCCGGCGCCCACTGCTGACAGTGCATATCCGAGGGGCTCACCTTCGAGCCGT 240
 Db 861 TTCCCGTATCCAGAGACCAACCGCGCGCCCTTACCAAGGAGGGCGCACCTTCGAGGCCG 920
 QY 241 GGTGGACACCGGTACAACACTTCAAGAGCTGGCGGCGCCCAACCCCAATCCCGGCTAT 300
 Db 921 GACCGCACCGGTACAACACTTCAAGAGCGCGACGCGCCCGCCCGATCCCGGCTAT 980
 QY 301 GCGGAGTAGTATCAAGAGCCAGTGTATGGCATAATATGCTACAGAGGTGTATACCT 360
 Db 981 GCGGAGTAGTATCAAGAGCCAGTGTATGGCATAATATGCTACAGAGGTGTATACCT 1037

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2279)
AUTHORS Shibata,H., Huynh,D.P., Vo,T.T. and Pulsat,S.-M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1998) Division of Neurology and Neurogenetics
Laboratory, Burns and Allen Research Institute, Cedars-Sinai
Medical Center, UCLA School of Medicine, 8700 Beverly Blvd., Los
Angeles, CA 90048, USA

FEATURES
Source
1..2279
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="similar to Homo sapiens CDNA clone HSP38A20"
1..2279
/gene="A2BP"
1047..2180
/gene="A2BP"
/codon_start=1
/product="ataxin-2-binding protein"
/protein_id="AAF78291.1"
/db_xref="GI:8671586"
/translation="MAQPYASAGPAPPPONGIPAEYTAHPHPAPRYTQTTPEHTLN
LYPAQTSQSPADTSAQTVSGATQTDPAPTDGPQOTQSENTENKSPKRLHVS
NIPFRFDPDLROMGQFGKILVDEIINERSGKGFVTESSADARBEKLGTV
VEGRKIEVNNAATVMTNKTVPYTNKMLNPVGAIVSEFYAGTVLLCOANQES
SMYSAPSLVYTSAMGFPYPAATAAAYRGALHGRGRTVNTFRAPAPPPIYAG
GVVODGPGADIVGVAAYRAOPTPAATAAAYSDRSYGVVADPHHALAPAPYGV
GMAAFAPLTDKTRSHADVGLVLSIQASLYRGYNRAFY"
1341..1358
/gene="A2BP"
/note="Region: RNA binding motif RNP-2"
1452..1472
/gene="A2BP"
/note="Region: RNA binding motif RNP-1"

BASE COUNT 497 a 712 c 636 g 434 t
ORIGIN

Query Match 82.0%; Score 499.2; DB 9; Length 2279;
Best Local Similarity 90.0%; Pred. No. 6.4e-134;
Matches 547; Conservative 0; Mismatches 58; Indels 3; Gaps 1;

QY 1 ATGACTAATAAAAGGCGGTGAACCCCTACACCAATGGCTGAAATTAATCCAGTTGTG 60
DB 1572 ATGACAAATAAAAAGACCGTCAACCTTATACAAATGGCTGAAATTAATCCAGTTGTG 1631
QY 61 GGGCGGCTCTACAGCCCCGACCTTCTATGACAGGACGGTGTGTTGGCCAGGCCAACG 120
DB 1632 GGTGCACTCTACAGTCCCGAATTTCTATGACAGGACGGTCTGTTGTGCCAGGCCAACG 1691
QY 121 GAGGAGTCTTCACTGATGAGTGGCCAGTTCACTTGATATATCTTGCAATGCTCTGGC 180
DB 1692 GAGGAGTCTTCACTGATGAGTGGCCAGTTCACTTGATATATCTTGCAATGCTCTGGC 1751
QY 181 TTTCATATCCGCGCCGCACTGCTGACGTGATACGAGGGGCTCACTTCAGAGCCGT 240
DB 1752 TTCCCGTATCCAGCACACCGCGGGGGCGCTCAAGAGGGCGACCTGCGAGGCCG 1811
QY 241 GGTGCAACCGGTGTACAACACTTCAAGAGTGGGGCGCCCCACCCCAATCCGGGCTAT 300
DB 1812 GGTGCAACCGGTGTACAACACTTCAAGAGTGGGGCGCCCCACCCCAATCCGGGCTAT 1871
QY 301 GGGGAGTAGTGATCAAGAGCAGGTATGGAATTAATGTCTACAGGGTGGTTACGCT 360
DB 1872 GGGGAGTAGTGATCAAGAGTGGTATTTATGCTTTCAGACACTTATATGATGCTTATGCT 1928
QY 361 GCATACCGCTACGCCAGCCACCTGACCACTGCTGCTGCTCAAGTACAGTTACGGA 420
DB 1929 GCATACCGCTACGCCAGCCACCTGACCACTGCTGCTGCTCAAGTACAGTTACGGA 1988
QY 421 CGAGTTATGCTGCCAGCCCTTACCAACAACACTTCTCAAGCCCCCAGCTTACAGGCTT 480
DB 1989 CGAGTTATGCTGCCAGCCCTTACCAACAACACTTCTCAAGCCCCCAGCTTACAGGCTT 2048

QY 481 GGTGCATGAATGCTTTTGGCCCTTGACCGATGCCAAGACTAGAGCCATGCTGATAT 540
DB 2049 GGTGCATGAATGCTTTTGGCCCTTGACCGATGCCAAGACTAGAGCCATGCTGATAT 2108
QY 541 GTGGCTCGCTTTCTTTCTTCAATGAGGCTAGTATATACCAAGGGGATACACCGCTT 600
DB 2109 GTGGCTCGCTTTCTTTCTTCAATGAGGCTAGTATATACCAAGGGGATACACCGCTT 2168
QY 601 GCTCCATA 608
DB 2169 GCTCCATA 2176

RESULT 8
AF109106
LOCUS AF109106
DEFINITION Homo sapiens hexaribonucleotide binding protein 1 isoform beta
ACCESSION AF109106
VERSION AF109106
KEYWORDS AF109106.1 GI:19032367

SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Chen,W., Chu,Z.-L., Blough,R.I., Liu,L., Hoppe,B. and
Winkelmann,J.C.
TITLE Molecular Cloning and Chromosomal Localization of a Novel Human
Brain, Heart and Skeletal Muscle Specific RNA Binding Protein Gene
Homologous to fox-1 in Caenorhabditis elegans
Unpublished
2 (bases 1 to 3348)

JOURNAL
REFERENCE Chen,W., Chu,Z.-L., Blough,R.I., Liu,L., Hoppe,B. and
Winkelmann,J.C.
TITLE Direct Submission
JOURNAL Submitted (24-NOV-1998) Internal Medicine/Hematology-Oncology,
University of Cincinnati College of Medicine, 231 Bethesda Ave. ML
0508, Cincinnati, OH 45267-0508, USA

FEATURES
Source
1..3348
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p13.3"
1..3348
/gene="HRNP1"
254..1441
/gene="HRNP1"
/note="HRNP1"
/note="similar to Caenorhabditis elegans fox-1"
/codon_start=1
/product="hexaribonucleotide binding protein 1 isoform
beta"
/protein_id="AAL83406.1"
/db_xref="GI:19032368"
/translation="MLASQGVLLHPYGVPMIVPAAPYLPGIIGNQAAAAADPTAQP
/translaction="MLASQGVLLHPYGVPMIVPAAPYLPGIIGNQAAAAADPTAQP
OTVSGTATQTDPAPTDGPQOTQSENTENKSPKRLHVSNIIPFRFDPDLROMGOF
GRILDELITENERSKGFVTFENSADADAREKHETVVEGRKIEVNNAATVMTN
KKTVPYTNKMLNPVGAIVSEFYAGTVLLCOANQESSMYSAPSLVYTSAMGFP
PYPAATAAAYRGALHGRGRTVNTFRAPAPPPIYAGGVVADPHHALAPAPYGV
AAVRAOPTPAATAAAYSDRQFVFAADEISCNISAVDEEMLPPTTTHLQPPPTA
LVP"
3299..3304
/gene="HRNP1"

BASE COUNT 950 a 775 c 711 g 912 t
ORIGIN

Query Match 78.7%; Score 479.4; DB 9; Length 3348;
Best Local Similarity 85.8%; Pred. No. 3.7e-128;
Matches 567; Conservative 0; Mismatches 41; Indels 53; Gaps 1;

QY 361 GCATACCGCTACGCCAGCCACCCCTGCTGCTGCTACAGTACAGTACCGA 420
|||
Db 1576 GCGTACCGCTATGCGCCAGCCACCCCTGCTGCTGCTACAGTACAGTACCGA 1635
|||
QY 421 CGAGTTATGCTGCGGACCCCTACACACACTTGTCCAGCCCCCAGCTACGGCGTT 480
|||
Db 1636 CGAGTTATGCTGCGGACCCCTACACACACTTGTCCAGCCCCCAGCTACGGCGTT 1695
|||
QY 481 GGTCCCATGATGCTTTTGGCCCTTGACCGATGCGCAAGACTAGAGCCATGCTGAT 540
|||
Db 1696 GGTCCCATGATGCTTTTGGCCCTTGACCGATGCGCAAGACTAGAGCCATGCTGAT 1755
|||
QY 541 GTGGGCTGCTTTCTTTCTTCACTTGGAGGCTAGTATATACCAAGGGGATACACCGTTT 600
|||
Db 1756 GTGGGCTGCTTTCTTTCTTCACTTGGAGGCTAGTATATACCAAGGGGATACACCGTTT 1815
|||
QY 601 GCTCCATAT 609
|||
Db 1816 GCTCCATAT 1824
|||
RESULT 6
AK001027 1513 bp mRNA linear PRI 01-AUG-2002
LOCUS Homo sapiens CDNA FLJ10165 f18, clone HEMBA1003591, weakly similar
DEFINITION to CHLOROPLAST 28 KD RIBONUCLEOPROTEIN PRECURSOR.
ACCESSION AK001027
VERSION AK001027.1 GI:7022045
KEYWORDS oligo capping, f18 (full insert sequence).
SOURCE Homo sapiens embryo, 10 weeks whole embryo, mainly head CDNA to
mRNA, clone 11b:HEMBA1 clone:HEMBA1003591.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
Isegai,T., Oca,T., Hayaeshi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
Nagatsuma,M., Hosokiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Matanabe,S., Kimura,K., Murakami,K.,
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
Nakamura,Y., Nagahara,K., Maehuo,Y., Nishimura,K. and Iwayanagi,T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 1513)
TITLE Isegai,T. and Otsuki,T.
JOURNAL Direct Submission
AUTHORS Submitted (16-FEB-2000) Takao Isegai, Helix Research Institute,
Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan, cDNA full insert
sequencing: Research Association for Biotechnology: cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
SOURCE Location/Qualifiers
1. .1513
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1003591"
/issue_type="whole embryo, mainly head"
/clone_lib="HEMBA1"
/dev_stage="embryo, 10 weeks"
/note="Cloning vector: PME18SFL3"
221. .1414
/notes="unamed protein product"
/codon_start=1
/protein_id="BAA91472.1"
/db_xref="GI:7022046"
/translation="MNCEREDLRGNOBAAAPDTMAQPYASAGAPFONGIPAEYTA

HPHPAEYTGQTTVPENTLNLTPPACTHSESPADPSAQTSGTATQADAPPDGP
OTQPSVENTKSDPKRLHVSNIPIPRFEDPLRQFGKILVEIIPNERSKGRF
VTEPNSADNRAREKLHGVNEGRKI EVNNAKTARWMTKTVPTNGLNVEYGAIV
YSEEFYAGTILCOANOQSGSNTSASSLVTYDAMPFPAPATAAAYGAILRGIG
RTVYNTFRRAAPPEPIPAYGVVYODGFADLYGGYAAARYAOPFPATAAASDSYG
RYVAADPYHHALAPAPLYGVGAMNAFAPLIDATRSHADVGLVLSLQASIRGVGN
RPAPY"
BASE COUNT 354 a 418 c 395 g 346 t
ORIGIN
Query Match 82.0%; Score 499.2; DB 9; Length 1513;
Best Local Similarity 90.0%; Pred. No. 6,1e-134;
Matches 547; Conservative 0; Mismatches 58; Indels 3; Gaps 1;
QY 1 ATGACTATATAAAGGCGGTGAGCCCTTACACCAATGCTGGAATTAATCCAGTTGTG 60
|||
Db 806 ATGACAATAAAGAACCGTCAACCTTATACAAATGGCTGGAATTAATCCAGTTGTG 865
|||
QY 61 GCGCGGCTCTACAGCCCGGACTTTCTATGACGACCGGTGCTGTGTGCGAGGCCAACG 120
|||
Db 866 GGTGAGCTCTACAGTCCCGAATTTCTATGACGACCGGTGCTGTGTGCGAGGCCAACG 925
|||
QY 121 GAGGATCTTCATGATACAGTGGGCGCCAGTTCATCTGTATATCTTGCAATGCTGGC 180
|||
Db 926 GAGGATCTTCATGATACAGTGGGCGCCAGTTCATCTGTATATCTTGCAATGCTGGC 985
|||
QY 181 TTTCATATCCGCGCGGCTGCTGACGCTGATACGAGGGGCTCACTTTCGAGCGCT 240
|||
Db 986 TTCCCGTATCCAGACGACGACCGCGCGCGGCTTACCGAGGGGCGACCTGCGAGGCGCG 1045
|||
QY 241 GGTGCGACCGTGTCAACACCTTCAAGAGCTGGGGGCGCCCAACCCCAATCCGGCTAT 300
|||
Db 1046 GGTGCGACCGTGTCAACACCTTCAAGAGCGCGCGCGCGCGCCCGCCCGATCCGCGCTAC 1105
|||
QY 301 GCGGAGAGTGTATCAAGAGCCAGTGTGCAATTAATGCTACAGGGGTGTACGCT 360
|||
Db 1106 GCGGAGTGTGTATCAAGAGTGTGCAATTAATGCTACAGGGGTGTACGCT 1162
|||
QY 361 GCATACCGCTACGCCAGCCACCCCTGCTGCTGCTACAGTACAGTACCGA 420
|||
Db 1163 GCATACCGCTACGCCAGCCACCCCTGCTGCTGCTACAGTACAGTACCGA 1222
|||
QY 421 CGAGTTATGCTGCGGACCCCTTACACACACTTGTCTCAGCCGCCACTTACGGCGTT 480
|||
Db 1223 CGAGTTATGCTGCGGACCCCTTACACACACTTGTCTCAGCCGCCACTTACGGCGTT 1282
|||
QY 481 GGTCCCATGATGCTTTTGGCCCTTGACCGATGCGCAAGACTAGAGCCATGCTGAT 540
|||
Db 1283 GGTCCCATGATGCTTTTGGCCCTTGACCGATGCGCAAGACTAGAGCCATGCTGAT 1342
|||
QY 541 GTGGGCTGCTTTCTTTCTTCACTTGGAGGCTAGTATATACCAAGGGGATACACCGTTT 600
|||
Db 1343 GTGGGCTGCTTTCTTTCTTCACTTGGAGGCTAGTATATACCAAGGGGATACACCGTTT 1402
|||
QY 601 GCTCCATAT 608
|||
Db 1403 GCTCCATAT 1410
|||
RESULT 7
AF107203 2279 bp mRNA linear PRI 23-JUN-2000
LOCUS Homo sapiens etaxin 2-binding protein (A2BP) mRNA, complete cds.
DEFINITION AF107203
ACCESSION AF107203
VERSION AF107203.1 GI:8671585
KEYWORDS
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2279)
AUTHORS Shibata,H., Huyenh,D.P., Vo,T.T. and Pulst,S.-M.
TITLE A novel protein, A2BP, with RNA binding motif binds to C-terminal

ORIGIN

Query Match 88.4%; Score 538.6; DB 10; Length 1363;
 Best Local Similarity 93.9%; Pred. No. 2.2e-145;
 Matches 572; Conservative 0; Mismatches 34; Indels 3; Gaps 1;

QY 1 ATGACTAATAAAGAGCCGCTGAACCCCTACACCAATGGCTGGAATTAATCACTGTTG 60
 DB 595 ATGACAAATTAAGAGAGCTGTCAACCCCTACACCAATGGCTGGAATTAATCACTGTTG 654
 QY 61 GCGCGGCTTACAGCCCGGCTCTATGAGGACGAGTCTGTTGGCCAGCCCAACCG 120
 DB 655 GCGCGGCTTACAGCCCGGCTCTATGAGGACGAGTCTGTTGGCCAGCCCAACCG 714
 QY 121 GAGGATCTTCATGATGACAGTGGCCCACTTCACTGTATATCTTCTGCAATGGCTGGC 180
 DB 715 GAGGATCTTCATGATGACAGTGGCCCACTTCACTGTATATCTTCTGCAATGGCTGGC 774
 QY 181 TTTCATATCCGCGCGCCCACTGCTGAGCTGATACCAAGGAGGCTCACCTTGAAGCCGT 240
 DB 775 TTCCATATCCGCGCGCCCACTGCTGAGCTGATACCAAGGAGGCTCACCTTGAAGCCGT 834
 QY 241 GGTGCGACCGGTATACCAACCTTCAAGCTGGCGGCGCCCAACCCCAATCCCGGCTAT 300
 DB 835 GGTGCGACCGGTATACCAACCTTCAAGCTGGCGGCGCCCAACCCCAATCCCGGCTAT 894
 QY 301 GCGGAGTAGTGTATCAAGAGCCAGTGTATGGCAATTAATGCTACAGGGGTGTTACGCT 360
 DB 895 GCGGAGTAGTGTATCAAGAGCCAGTGTATGGCAATTAATGCTACAGGGGTGTTACGCT 951
 QY 361 GCATACCGCTACGCGCCAGCCCAACCCCTGCACTGCTGCTGCTACAGTACAGTTACGGA 420
 DB 952 GCGTACCGCTATGCGCCAGCCCAACCCCTGCACTGCGCTGCTGCTACAGTACAGTTACGGA 1011
 QY 421 CGAGTTATGCTGCGGACCCCTTACCAACCACTTGTCTCCAGCCCCCACTTACGCGCTT 480
 DB 1012 CGAGTTATGCTGCGGACCCCTTACCAACCACTTGTCTCCAGCCCCCACTTACGCGCTT 1071
 QY 481 GGTGCGACCAATGCTTTGGCGGCTTACAGGAGTGGCAAGTACAGGAGCATGCTGAT 540
 DB 1072 GGTGCGACCAATGCTTTGGCGGCTTACAGGAGTGGCAAGTACAGGAGCATGCTGAT 1131
 QY 541 GTGGGCTGCTCTTCTTCTTCACTTGCAGGCTAGTATATACCAAGGAGGATACCAACGTTT 600
 DB 1132 GTGGGCTGCTCTTCTTCTTCACTTGCAGGCTAGTATATACCAAGGAGGATACCAACGTTT 1191
 QY 601 GCTCCATAT 609
 DB 1192 GCTCCATAT 1200

RESULT 5
 AB041596 2002 bp mRNA linear ROD 30-JUN-2000
 LOCUS Mus musculus brain cDNA, clone MNCD-3035, similar to Homo sapiens
 DEFINITION cDNA FLJ10165, clone HEMBA1003591.
 ACCESSION AB041596
 VERSION AB041596.1 GI:7670455
 KEYWORDS f1s (full insert sequence).
 SOURCE Mus musculus (strain:C57BL) adult female cDNA to mRNA,
 clone 11b:Sugano mouse brain mncb clone:MNCB-3035.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (sites)
 AUTHORS Osada, N., Kusuda, J., Tanuma, R., Ito, A., Hirata, M., Sugano, S. and Hashimoto, K.
 TITLE Isolation of full-length cDNA clones from mouse brain cDNA library made by oligo-capping method
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2002)
 AUTHORS Hashimoto, K., Osada, N., Kusuda, J. and Sugano, S.
 TITLE Direct Submission

JOURNAL

Submitted (12-APR-2000) Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
 (E-mail: khashi@nih.go.jp, URL: <http://www.nih.go.jp/yoken/genebank/>,
 Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)
 URL: <http://www.nih.go.jp/yoken/genebank/>

COMMENT

Lab Name: Sugano mouse brain mncb
 Lab host: TOPI0
 Vector: PME18S-FLJ
 1st strand cDNA was primed with an oligo(dT) primer
 [ATGTCGCTTTTCTTTTCTTTT]; double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of PME18S-FLJ. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Library was constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing (5' end primer [CTTGTGCTCTAAAGCTGCG]; 3' end primer [CGACCTGCACTGAGACAC]).
 A part of this sequence is reported in AU067167.
 Location/Qualifiers

FEATURES

source
 1..2002
 /organism="Mus musculus"
 /strain="C57BL"
 /db_xref="taxon:10090"
 /clone="MNCD-3035"
 /sex="female"
 /clone_lib="Sugano mouse brain mncb"
 /dev_stage="adult"
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BA95079.1"
 /db_xref="GI:7670455"
 /translation="MNCERQLRNOEMAAAPDTMAOPYASQFAPRONGIPAEYAP
 HPHAPRYGQTVYGHLLNLYPTQHSSESAITSQTVSGIATQDDAPTDGQPO
 TOPSEBESKSPQRLHVSNI PFRFPDLMOMFGORSLDVEII FENERSKQGFV
 TFENSADARAREKLHGTVVEGRKI EVNNAFARVMTNKTVNPTNGKLPVVGAVY
 SPDYAGTVLLCANOGSSMSYSGPSSILVTSAMPFPYPAATAAARGLALNSGR
 TVYVTPRAAAPPRPIPAYGCVYVDGSGADYIGGVAAYRAQDTPATPAAYASQYGR
 VTADPYHTLPAFTPTTGVMNMFAPLTDKTRGHADVDVLSSLOASLYRGRYNR
 PAPY"

CDS

BASE COUNT 518 a 589 c 516 g 379 t
 ORIGIN
 Query Match 88.4%; Score 538.6; DB 10; Length 2002;
 Best Local Similarity 93.9%; Pred. No. 2.2e-145;
 Matches 572; Conservative 0; Mismatches 34; Indels 3; Gaps 1;

QY 1 ATGACTAATAAAGAGCCGCTGAACCCCTACACCAATGGCTGGAATTAATCACTGTTG 60
 DB 1219 ATGACAAATTAAGAGAGCTGTCAACCCCTACACCAATGGCTGGAATTAATCACTGTTG 1278
 QY 61 GCGCGGCTTACAGCCCGGCTCTATGAGGACGAGTCTGTTGGCCAGCCCAACCG 120
 DB 1279 GCGCGGCTTACAGCCCGGCTCTATGAGGACGAGTCTGTTGGCCAGCCCAACCG 1338
 QY 121 GAGGATCTTCATGATGACAGTGGCCCACTTCACTGTATATCTTCTGCAATGGCTGGC 180
 DB 1339 GAGGATCTTCATGATGACAGTGGCCCACTTCACTGTATATCTTCTGCAATGGCTGGC 1398
 QY 181 TTTCATATCCGCGCGCCCACTGCTGAGCTGATACCAAGGAGGCTCACCTTGAAGCCGT 240
 DB 1399 TTCCATATCCGCGCGCCCACTGCTGAGCTGATACCAAGGAGGCTCACCTTGAAGCCGT 1458
 QY 241 GGTGCGACCGGTATACCAACCTTCAAGCTGGCGGCGCCCAACCCCAATCCCGGCTAT 300
 DB 1459 GGTGCGACCGGTATACCAACCTTCAAGCTGGCGGCGCCCAACCCCAATCCCGGCTAT 1518
 QY 301 GCGGAGTAGTGTATCAAGAGCCAGTGTATGGCAATTAATGCTACAGGGGTGTTACGCT 360
 DB 1519 GCGGAGTAGTGTATCAAGAGCCAGTGTATATGG--TGACACATTTATGTTATGCTT 1575

Db	1408	GCTCCATAT	1416
RESULT 3	AF229057	2000 bp	mRNA linear PRI 01-FEB-2002
LOCUS	AF229057		
DEFINITION	Homo sapiens hexaribonucleotide binding protein 1 isoform gamma (HRNBPI) mRNA, complete cds.		
ACCESSION	AF229057		
VERSION	AF229057.1	GI:18461366	
KEYWORDS			
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 2000)		
TITLE	Chen, W. and Winkelman, J.C.		
JOURNAL	Direct Submission		
FEATURES	Submitted (28-JUN-2000) Internal Medicine/Hematology-Oncology, University of Cincinnati College of Medicine, The Vontz Center for Molecular Studies, 3125 Eden Avenue, Cincinnati, OH 45267-0508, USA		
source	Location/Qualifiers		
gene	1..2000		
CDS	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="16"		
	/map="16p13.3"		
	1..2000		
	/gene="HRNBPI"		
	256..1512		
	/gene="HRNBPI"		
	/note="RNA-binding protein"		
	/codon_start=1		
	/product="hexaribonucleotide binding protein 1 isoform gamma"		
	/protein_id="AAL71904.1"		
	/db_xref="GI:18461367"		
	/translation="MLASQVLLHPYGVNTPVAPYPIGLIGNQEMAAAPDTMAOPVASFAPPNQIGPAEYVAPHPAPETGQTVPEPENTLNPYPAQTHSESPADTSAQTVGATOTDDAAPDQOPQOTPOSENTEMSQPRRLVNSI PERFPDRLRMFGQNGKILDEVEII FENRSGKGFVTPENSADBARERKLQTVVEGRKIEVNNAITAMTNRGKIDPYNPKGKLVNIVGASVPEYAGTVLLCOANOBSGMSYAPSILVYTSAMPGPPYPATTAARGAHLRGRGTVNTPFAALPPPIPYGVGVVYQEPYVNGKLLQGVNAAVYAOFTPATTAAYSDSYRVRVAADPYYHALPAPYVGVAMNAPFLDTAKTRSHADDVGLVSSIQASITVYRGYNRFAPY"		
BASE COUNT	502 a 505 c 487 g 506 t		
ORIGIN			
Query Match	89.1%	Score 542.4;	DB 9; Length 2000;
Best Local Similarity	93.3%	Pred. No. 1.8e-146;	
Matches 567; Conservative	0; Mismatches 41; Indels 0; Gaps 0;		
1	ATGACTAATTAAGGCGGTGAACCCCTTACACCAATGCGTGAATAATTCAGATTGG	60	
901	ATGACMAATTAAGGCGGTGAACCCCTTACACCAATGCGTGAATAATTCAGATTGG	960	
61	GCGCGGCTTACAGCCCGCATCTTATGACGAGCAGCGTCTGTGCGAGGCCAAG	120	
961	GGTGACGTCTACAGTCCGGAATTTATATGACGAGCAGCGTCTGTGCGAGGCCAAG	1020	
121	GAGGATCTTTCATGTACAGTGGCCCGCATCTTATGACGAGCAGCGTCTGTGCGAGGCCAAG	180	
1021	GAGGATCTTTCATGTACAGTGGCCCGCATCTTATGACGAGCAGCGTCTGTGCGAGGCCAAG	1080	
181	TTTCATATCCGCGCGCATCTGTCAGTGCATCCGAGGGGCTCACCTTGAGAGCGGT	240	
1081	TTTCATATCCGCGCGCATCTGTCAGTGCATCCGAGGGGCTCACCTTGAGAGCGGT	1140	
241	GGTGACCGGTGACACCACTTTCAGAGCTGCGGCGCCCGCATCCCGGCTAT	300	
1141	GGTGACCGGTGACACCACTTTCAGAGCGTGGGCGGCGCCCGCATCCCGGCTAT	1200	
301	GCGGAGTATGATATCAGACCGCAGTGTATGGCAATAATTCCTACAGGAGTGTACGCT	360	

Db	1201	GCCGAGTAGTGTATCAGAGAGCTGTGTATGCAATTAATTGCTGCAGAGTGTATTGCT	1260
QY	361	GCATACCGCTACGCGCCAGCCACCCTCTGCACCTGCTGCTGCTTACAGTACGATTACGGA	420
Db	1261	GCATACCGCTACGCGCCAGCCACCCTCTGCACCTGCTGCTGCTTACAGTACGATTACGGA	1320
QY	421	CGAGTTTATGCTGCCGACCCCTTACCAACACACTTGTCCAGCCCCCACTACGGCGTT	480
Db	1321	CGAGTTTATGCTGCCGACCCCTTACCAACACACTTGTCCAGCCCCCACTACGGCGTT	1380
QY	481	GATGCCATGAATGCTTTTGGCGCCCTTGACCGATGCAGCAAGATCAGAGCATGTGATGAT	540
Db	1381	GATGCCATGAATGCTTTTGGCGCCCTTGACCGATGTGACATGCAAGATCAGAGCATGTGATGAT	1440
QY	541	GTGGGCTCTGTTCTTTCTTTCTTCAATGCAAGGCTAGTATATACCAAGGGGGATACACCGTTT	600
Db	1441	GTGGGCTCTGTTCTTTCTTTCTTCAATGCAAGGCTAGTATATACCAAGGGGGATACACCGTTT	1500
QY	601	GCTCCATA 608	
Db	1501	GCTCCATA 1508	
RESULT 4			
AF107204			
LOCUS	AF107204	1363 bp	mRNA linear ROD 23-JUN-2000
DEFINITION	Mus musculus ataxin 2-binding protein (A2bp) mRNA, complete cds.		
ACCESSION	AF107204		
VERSION	AF107204.1	GI:8671587	
KEYWORDS			
SOURCE	Mus musculus.		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 1363)		
AUTHORS	Shibata,H., Huynh,D.P., Vo,T.T. and Pulst,S.-M.		
TITLE	A novel protein, A2BP, with RNA binding motif, binds to C-terminal ataxin-2		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 1363)		
AUTHORS	Shibata,H., Huynh,D.P., Vo,T.T. and Pulst,S.-M.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-NOV-1998) Division of Neurology and Neurogenetics Laboratory, Burns and Allen Research Institute, Cedars-Sinai Medical Center, UCLA School of Medicine, 8700 Beverly Blvd., Los Angeles, CA 90048, USA		
FEATURES			
source	1..1363		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
gene	1..1363		
	/gene="A2bp"		
CDS	67..1203		
	/gene="A2bp"		
	/codon_start=1		
	/product="ataxin 2-binding protein"		
	/protein_id="AA78292.1"		
	/db_xref="GI:8671588"		
	/translation="MAQPVSAQFAPRONGIPAEYTAHPHPAPAEYGTTPDHTLNLYPTQTHSOSADTASQVSTGATRTDNDAPTDGQOQPSNTSKSQPKRLHVNIPPRFRDPDLDMRQGFQFKIDVLEITPERSKSGGFYTFPNSADADAPREKLTHTVGRKRLVNNATARWQNTKNTVNPYTNKLNLPVGVAVTSPDYAGTLLCQANDQSGSVNSGSSSLVYTSAMGFPYPAATAAARGLRGRVTYNTFPAALPPEIPIYVGVVODGFYGADIVGSGYAAVRYAQTPATAAALVSDSYGRVYADAPVHHTLAPPTGVGAMNNAPLTDAKTRSHADVGLVLSLQASIVRGYNFAPY"		
misc_feature	358..375		
	/gene="A2bp"		
misc_feature	475..495		
	/note="Region: RNA binding motif RNP-2"		
misc_feature	475..495		
	/gene="A2bp"		
	/note="Region: RNA binding motif RNP-1"		
BASE COUNT	362 a 396 c 324 g 281 t		

JOURNAL Patent: WO 0174901-A 1 11-OCT-2001;
Scios Inc. (US)
Location/Qualifiers
FEATURES 1. 1340
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
BASE COUNT 366 a 394 c 317 g 263 t
ORIGIN

Query Match 100.0%; Score 609; DB 6; Length 1340;
Best Local Similarity 100.0%; Pred. No. 7.5e-166;
Matches 609; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACTAATAAAGAGCCGCTGAACCCCTACACCAATGCTGGAATTAATCACTGTTG 60
DB 535 ATGACTAATAAAGAGCCGCTGAACCCCTACACCAATGCTGGAATTAATCACTGTTG 594
QY 61 GCGCGGTCTACAGCCCGCACTTTATGCAAGCAGGTCTGTTGTGCCAGGCCAACAG 120
DB 595 GCGCGGTCTACAGCCCGCACTTTATGCAAGCAGGTCTGTTGTGCCAGGCCAACAG 654
QY 121 GAGGATCTTCATGATGAGTGGCCCACTTGTATATCTTGTCAATGCTGGC 180
DB 655 GAGGATCTTCATGATGAGTGGCCCACTTGTATATCTTGTCAATGCTGGC 714
QY 181 TTTCATATCCGCGCCGCACTGCTGAGCTGATACCGAGGGCTCACCTTGCAGGCGT 240
DB 715 TTTCATATCCGCGCCGCACTGCTGAGCTGATACCGAGGGCTCACCTTGCAGGCGT 774
QY 241 GGTGCGACCGGTACCAACCTTTACAGTGGCGGCGCCCAACCCCAATCCCGGCTAT 300
DB 775 GGTGCGACCGGTACCAACCTTTACAGTGGCGGCGCCCAACCCCAATCCCGGCTAT 834
QY 301 GCGCGAGTGTATCAAGCAGGTATGGAATTAATGCTAAGGCTGTTACGCT 360
DB 835 GCGCGAGTGTATCAAGCAGGTATGGAATTAATGCTAAGGCTGTTACGCT 894
QY 361 GATACCGCTACGCGCCAGCCACCCCTGCACTGCTGCTACAGTACAGTACGGA 420
DB 895 GATACCGCTACGCGCCAGCCACCCCTGCACTGCTGCTACAGTACAGTACGGA 954
QY 421 CGAGTTATGCTGCGCAGCCCTTACACCAACACTTGTCTCCAGCCCACTTACGCGCTT 480
DB 955 CGAGTTATGCTGCGCAGCCCTTACACCAACACTTGTCTCCAGCCCACTTACGCGCTT 1014
QY 481 GGTGCGATGAATGCTTTTGGCCCTTACAGCAGTCCCAAGTGAAGTGTATGAT 540
DB 1015 GGTGCGATGAATGCTTTTGGCCCTTACAGCAGTCCCAAGTGAAGTGTATGAT 1074
QY 541 GTGGGCTGCTCTTCTTCTCATGAGGCTAGTATACCAAGGGGATACCAACGTTT 600
DB 1075 GTGGGCTGCTCTTCTTCTCATGAGGCTAGTATACCAAGGGGATACCAACGTTT 1134
QY 601 GGTCCATAT 609
DB 1135 GGTCCATAT 1143

RESULT 2
AF191501 1586 bp mRNA linear ROD 01-MAR-2002
LOCUS AF191501
DEFINITION Mus musculus hexaribonucleotide binding protein 1 (Hrbp1) mRNA,
complete cds.
ACCESSION AF191501 GI:19032413
VERSION AF191501.1
KEYWORDS
SOURCE
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1586)
AUTHORS Chen, W. and Winkelman, J. C.
TITLE Direct Submission

JOURNAL Submitted (01-OCT-1999) Internal Medicine/Hematology-Oncology,
University of Cincinnati College of Medicine, 231 Bethesda Ave.,
Cincinnati, OH 45267-0508, USA
Location/Qualifiers
FEATURES 1. 1586
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="skeletal muscle"
BASE COUNT 407 a 458 c 374 g 347 t
ORIGIN

Query Match 95.3%; Score 580.2; DB 10; Length 1586;
Best Local Similarity 97.0%; Pred. No. 1.8e-157;
Matches 591; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 ATGACTAATAAAGAGCCGCTGAACCCCTACACCAATGCTGGAATTAATCACTGTTG 60
DB 808 ATGACTAATAAAGAGCCGCTGAACCCCTACACCAATGCTGGAATTAATCACTGTTG 867
QY 61 GCGCGGTCTACAGCCCGCACTTTATGCAAGCAGGTCTGTTGTGCCAGGCCAACAG 120
DB 868 GCGCGGTCTACAGCCCGCACTTTATGCAAGCAGGTCTGTTGTGCCAGGCCAACAG 927
QY 121 GAGGATCTTCATGATGAGTGGCCCACTTGTATATCTTGTCAATGCTGGC 180
DB 928 GAGGATCTTCATGATGAGTGGCCCACTTGTATATCTTGTCAATGCTGGC 987
QY 181 TTTCATATCCGCGCCGCACTGCTGAGCTGATACCGAGGGCTCACCTTGAAGCGCT 240
DB 988 TTTCATATCCGCGCCGCACTGCTGAGCTGATACCGAGGGCTCACCTTGAAGCGCT 1047
QY 241 GGTGCGACCGGTACCAACCTTTACAGTGGCGGCGCCCAACCCCAATCCCGGCTAT 300
DB 1048 GGTGCGACCGGTACCAACCTTTACAGTGGCGGCGCCCAACCCCAATCCCGGCTAT 1107
QY 301 GGTGCGATGAATGCTTTTGGCCCTTACAGCAGTCCCAAGTGAAGTGTATGAT 360
DB 1108 GGTGCGATGAATGCTTTTGGCCCTTACAGCAGTCCCAAGTGAAGTGTATGAT 1167
QY 361 GATACCGCTACGCGCCAGCCACCCCTGCACTGCTGCTACAGTACAGTACGGA 420
DB 1168 GATACCGCTACGCGCCAGCCACCCCTGCACTGCTGCTACAGTACAGTACGGA 1227
QY 421 CGAGTTATGCTGCGCAGCCCTTACACCAACACTTGTCTCCAGCCCACTTACGCGCTT 480
DB 1228 CGAGTTATGCTGCGCAGCCCTTACACCAACACTTGTCTCCAGCCCACTTACGCGCTT 1287
QY 481 GGTGCGATGAATGCTTTTGGCCCTTACAGCAGTCCCAAGTGAAGTGTATGAT 540
DB 1288 GGTGCGATGAATGCTTTTGGCCCTTACAGCAGTCCCAAGTGAAGTGTATGAT 1347
QY 541 GTGGGCTGCTCTTCTTCTCATGAGGCTAGTATACCAAGGGGATACCAACGTTT 600
DB 1348 GTGGGCTGCTCTTCTTCTCATGAGGCTAGTATACCAAGGGGATACCAACGTTT 1407
QY 601 GGTCCATAT 609

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 10:46:57 ; Search time 1628.58 Seconds

(without alignments)
10882.841 Million cell updates/sec

Title: US-09-809-545A-1_COPY_535_1143

Perfect score: 609
Sequence: 1 atgactaataaagagcgcgt.....acaacgcttcgctccacat 609Scoring table: IDENTITY NUC
Gapop 10*0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
GenEmbl.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pac.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_ets.*
12: gb_ey.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pac.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_ets.*
28: em_un.*
29: em_vl.*
30: em_hcg_hum.*
31: em_hcg_inv.*
32: em_hcg_other.*
33: em_hcg_mus.*
34: em_hcg_pln.*
35: em_hcg_rtd.*
36: em_hcg_mam.*
37: em_hcg_vrc.*
38: em_ay.*
39: em_hngo_hum.*
40: em_hngo_mus.*
41: em_hngo_other.*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	609	100.0	1340	6 AX268800	AX268800 Sequence
2	580.2	95.3	1586	10 AF191501	AF191501 Mus muscu
3	542.4	89.1	2000	9 AF229057	AF229057 Homo sapi
4	538.6	88.4	1363	10 AF107204	AF107204 Mus muscu
5	538.6	88.4	2002	10 AB041596	AB041596 Mus muscu
6	499.2	82.0	1513	9 AK001027	AK001027 Homo sapi
7	499.2	82.0	2279	9 AF107203	AF107203 Homo sapi
8	479.4	78.7	3348	9 AF109106	AF109106 Homo sapi
9	433	71.1	1475	9 AB060859	AB060859 Macaca fa
10	429.8	70.6	1547	9 AF094849	AF094849 Homo sapi
11	428.6	70.4	2372	6 AF134676	AF134676 Sequence
12	172.6	28.3	1623	9 HSM803066	AL173700 Homo sapi
13	119.4	19.6	115860	2 AC120661	AC120661 Rattus no
14	113.6	18.7	2971	9 HSFOX14	AF109120 Homo sapi
15	113.6	18.7	164538	9 AC005774	AC005774 Homo sapi
16	111	18.2	251	9 HSFOX09	AF109115 Homo sapi
17	87.4	14.4	396	9 HSA32368	AF229056 Mus muscu
18	87.4	14.4	617	9 HSA335811	AF229055 Mus muscu
19	87.4	14.4	653	9 HSA335616	AF229054 Mus muscu
20	87.4	14.4	675	9 HSA334013	AF229053 Mus muscu
21	81.8	13.4	374	9 HSFOX08	AF229052 Mus muscu
22	80.8	13.3	73929	2 AC103207	AF229051 Mus muscu
23	80.4	13.2	1134	10 AF387322	AF229050 Mus muscu
24	80.4	13.2	1695	10 AF229055	AF229049 Mus muscu
25	78.2	12.8	208632	2 AC073624	AF229048 Mus muscu
26	77.4	12.7	234	9 HSFOX13	AF229047 Mus muscu
27	77.2	12.7	209461	2 AC079359	AF229046 Mus muscu
28	75.6	12.4	1538	9 BC013115	AF229045 Mus muscu
29	73.2	12.0	959	10 AF229056	AF229044 Mus muscu
30	70.4	11.6	159565	2 AC1591075	AF229043 Mus muscu
31	69.8	11.5	115860	10 AL591075	AF229042 Mus muscu
32	66.8	11.0	1393	9 AY072786	AF229041 Mus muscu
33	66.8	11.0	1555	9 BC025281	AF229040 Mus muscu
34	66.8	11.0	1925	9 AF228058	AF229039 Mus muscu
35	66.4	10.9	1721	10 BC027263	AF229038 Mus muscu
36	66.2	10.9	151	9 HSFOX11	AF229037 Mus muscu
37	65.2	10.7	1558	9 AK055213	AF229036 Mus muscu
38	63.2	10.4	1876	9 HSP38A20	AF229035 Mus muscu
39	51.2	8.4	347	9 HSFOX07	AF229034 Mus muscu
40	51.2	8.4	112976	9 AC027683	AF229033 Mus muscu
41	48.2	7.9	156549	2 AC079072	AF229032 Mus muscu
42	45.6	7.5	125020	9 AF429315	AF229031 Mus muscu
43	45	7.4	2397	9 AK091695	AF229030 Mus muscu
44	44.6	7.3	428	6 AX336745	AF229029 Mus muscu
45	42	6.9	172090	2 AC018477	AF229028 Mus muscu

ALIGNMENTS

```
RESULT 1
AX268800
LOCUS AX268800 1340 bp DNA linear PAT 29-OCT-2001
DEFINITION Sequence 1 from Patent WO0174901.
ACCESSION AX268800
VERSION AX268800.1 GI:16541860
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 Stanton, L.W. and White, R.T.
AUTHORS
TITLE Secreted factors
```

RESULT 15
US-09-960-352-10167/C

; Sequence 10167, Application US/09960352
; Patent No. US20020137139A1

; GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.

; APPLICANT: Tao, Nengbing

; APPLICANT: Byatt, John C.

; APPLICANT: Mathialagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

; FILE REFERENCE: 16511.006/37-21(10298)C

; CURRENT APPLICATION NUMBER: US/09/960.352

; NUMBER OF SEQ ID NOS: 15112

; SEQ ID NO 10167

; LENGTH: 382

; TYPE: DNA

; ORGANISM: Bos taurus

; OTHER INFORMATION: Clone ID: 44-LIB188-014-Q1-E1-C8

US-09-960-352-10167

Query Match 2.0%; Score 27; DB 10; Length 382;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1294 AAAAAAAAAAAAAAAAAAATAGC 1320
|||
Db 31 AAAAAAAAAAAAAAAAAAATAGC 5

Search completed: March 16, 2003, 02:34:54
Job time : 127.631 secs

```

; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 714
; LENGTH: 308
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 04-LIB188-006-Q1-E1-A7
; US-09-960-352-714

Query Match
Best Local Similarity 100.0%; Score 27; DB 10; Length 308;
Pred. No. 0.18; Mismatches 0; Indels 0; Gaps 0;
Matches 27; Conservative 0;

QY 1294 AAAAAAAAAAAAAAAAAAATAGC 1320
DB 33 AAAAAAAAAAAAAAAAAAATAGC 7

RESULT 11
US-09-960-352-1236/c
; Sequence 1236, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 12366
; LENGTH: 315
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 53-LIB2809-032-Q1-E1-F2
; US-09-960-352-12366

Query Match
Best Local Similarity 100.0%; Score 27; DB 10; Length 315;
Pred. No. 0.18; Mismatches 0; Indels 0; Gaps 0;
Matches 27; Conservative 0;

QY 1294 AAAAAAAAAAAAAAAAAAATAGC 1320
DB 31 AAAAAAAAAAAAAAAAAAATAGC 5

RESULT 12
US-09-960-352-7832/c
; Sequence 7832, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 7832
; LENGTH: 337
; TYPE: DNA
```

```

; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 34-LIB188-006-Q1-E1-A6
; US-09-960-352-7832

Query Match
Best Local Similarity 100.0%; Score 27; DB 10; Length 337;
Pred. No. 0.17; Mismatches 0; Indels 0; Gaps 0;
Matches 27; Conservative 0;

QY 1294 AAAAAAAAAAAAAAAAAAATAGC 1320
DB 63 AAAAAAAAAAAAAAAAAAATAGC 37

RESULT 13
US-09-960-352-7833/c
; Sequence 7833, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 7833
; LENGTH: 344
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 34-LIB188-007-Q1-E1-A6
; US-09-960-352-7833

Query Match
Best Local Similarity 100.0%; Score 27; DB 10; Length 344;
Pred. No. 0.17; Mismatches 0; Indels 0; Gaps 0;
Matches 27; Conservative 0;

QY 1294 AAAAAAAAAAAAAAAAAAATAGC 1320
DB 31 AAAAAAAAAAAAAAAAAAATAGC 5

RESULT 14
US-09-960-352-11598/c
; Sequence 11598, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11598
; LENGTH: 349
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 50-LIB188-011-Q1-E1-E6
; US-09-960-352-11598

Query Match
Best Local Similarity 100.0%; Score 27; DB 10; Length 349;
Pred. No. 0.17; Mismatches 0; Indels 0; Gaps 0;
Matches 27; Conservative 0;

QY 1294 AAAAAAAAAAAAAAAAAAATAGC 1320
DB 31 AAAAAAAAAAAAAAAAAAATAGC 5
```


; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-42

Query Match 2.2%; Score 30; DB 10; Length 1574;
Best Local Similarity 100.0%; Pred. No. 0.0093;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1286 AACACGAAAAAAAAAAAAAAAAAAAA 1315
Db 1534 AACACGAAAAAAAAAAAAAAAAAAAA 1563

RESULT 6

US-09-960-352-289/C
; Sequence 289, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Machialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960.352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 289
; LENGTH: 289
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 02-LIB2809-015-Q1-E1-A5
US-09-960-352-289

Query Match 2.1%; Score 28; DB 10; Length 289;
Best Local Similarity 100.0%; Pred. No. 0.073;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1294 AAAAAAAAAAAAAAAAAATAAGCG 1321
Db 36 AAAAAAAAAAAAAAAAAATAAGCG 9

RESULT 7

US-09-867-701-9672
; Sequence 9672, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867.701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9672
; LENGTH: 177
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9672

Query Match 2.0%; Score 27; DB 10; Length 177;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1289 CAACGAAAAAAAAAAAAAAAAAAAA 1315
Db 133 CAACGAAAAAAAAAAAAAAAAAAAA 159

RESULT 8

US-09-983-965-591/C
; Sequence 591, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Machialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983.965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 591
; LENGTH: 214
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 02-LIB188-027-Q1-E1-A5
US-09-983-965-591

Query Match 2.0%; Score 27; DB 10; Length 214;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1294 AAAAAAAAAAAAAAAAAATAAGCG 1320
Db 44 AAAAAAAAAAAAAAAAAATAAGCG 18

RESULT 9

US-09-867-701-9696/C
; Sequence 9696, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867.701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9696
; LENGTH: 263
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9696

Query Match 2.0%; Score 27; DB 10; Length 263;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1289 CAACGAAAAAAAAAAAAAAAAAAAA 1315
Db 40 CAACGAAAAAAAAAAAAAAAAAAAA 14

RESULT 10

US-09-960-352-714/C
; Sequence 714, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing

PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 28375
LENGTH: 125
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC006075.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.67
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.48
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.57
OTHER INFORMATION: EST HUMAN HIT: Z25303.1, EVALUE 9.00e-61
OTHER INFORMATION: NT HIT: g11431054, EVALUE 2.00e-64
OTHER INFORMATION: SWISSPROT HIT: P54537, EVALUE 1.40e+00
US-09-864-761-28375

Query Match 2.4% Score 32; DB 10; Length 125;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 247 GGGCAGCCCGACACACACCTTCTGAAAAACAC 278
Db 9 GGGCAGCCCGACACACACCTTCTGAAAAACAC 40

RESULT 4
US-09-864-761-11804
Sequence 11804, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 11804
LENGTH: 473
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC006075.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.67
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.48
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.57
US-09-864-761-11804

Query Match 2.4% Score 32; DB 10; Length 473;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 247 GGGCAGCCCGACACACACCTTCTGAAAAACAC 278
Db 296 GGGCAGCCCGACACACACCTTCTGAAAAACAC 327

RESULT 5
US-09-925-302-42
Sequence 42, Application US/09925302
Patent No. US20020044941A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA104
CURRENT APPLICATION NUMBER: US/09/925,302
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05918
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 896
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 42
LENGTH: 1574
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1029)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (1076)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (1574)

GenCore version 5.1.4.D5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 16:12:39 ; Search time 119.631 Seconds
(without alignments)
7861.696 Million cell updates/sec

Title: US-09-809-545A-1
Perfect score: 1340
Sequence: 1 ggcggccgccccgcagacaat.....ggccgaagggtcgcctaga 1340

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 501302 seqs, 350932545 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1002604

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1340	100.0	1340	10	US-09-809-545A-1
2	70	5.2	2372	10	US-09-794-591-1
3	32	2.4	125	10	US-09-864-761-28375
4	32	2.4	473	10	US-09-864-761-11804
5	30	2.2	1574	10	US-09-925-302-42
6	28	2.1	289	10	US-09-860-352-289
7	27	2.0	177	10	US-09-867-701-9672
8	27	2.0	214	10	US-09-983-965-591
9	27	2.0	263	10	US-09-867-701-9696
10	27	2.0	308	10	US-09-960-352-714
11	27	2.0	315	10	US-09-960-352-12366
12	27	2.0	337	10	US-09-960-352-7832
13	27	2.0	344	10	US-09-960-352-7833
14	27	2.0	349	10	US-09-960-352-11598
15	27	2.0	382	10	US-09-960-352-10167
16	27	2.0	383	10	US-09-960-352-13506
17	27	2.0	388	10	US-09-960-352-6680
18	27	2.0	394	10	US-09-817-318-16
19	27	2.0	411	10	US-09-960-352-3551

C	20	27	2.0	418	10	US-09-960-352-10650	Sequence 10650, A
C	21	27	2.0	419	10	US-09-960-352-11041	Sequence 13041, A
C	22	27	2.0	434	10	US-09-960-352-12123	Sequence 12123, A
C	23	27	2.0	438	10	US-09-924-035A-516	Sequence 516, App
C	24	26	1.9	222	10	US-09-867-701-8979	Sequence 8979, App
C	25	26	1.9	230	10	US-09-920-300A-96	Sequence 96, App
C	26	26	1.9	230	10	US-09-920-300A-96	Sequence 96, App
C	27	26	1.9	230	10	US-09-920-300A-96	Sequence 96, App
C	28	26	1.9	230	10	US-09-920-300A-96	Sequence 96, App
C	29	26	1.9	230	10	US-09-920-300A-96	Sequence 96, App
C	30	26	1.9	230	10	US-09-920-300A-96	Sequence 96, App
C	31	26	1.9	230	10	US-09-920-300A-96	Sequence 96, App
C	32	26	1.9	230	10	US-09-920-300A-96	Sequence 96, App
C	33	26	1.9	230	10	US-09-920-300A-96	Sequence 96, App
C	34	26	1.9	230	10	US-09-920-300A-96	Sequence 96, App
C	35	26	1.9	230	10	US-09-920-300A-96	Sequence 96, App
C	36	26	1.9	230	10	US-09-920-300A-96	Sequence 96, App
C	37	26	1.9	230	10	US-09-920-300A-96	Sequence 96, App
C	38	26	1.9	230	10	US-09-920-300A-96	Sequence 96, App
C	39	26	1.9	230	10	US-09-920-300A-96	Sequence 96, App
C	40	26	1.9	230	10	US-09-920-300A-96	Sequence 96, App
C	41	26	1.9	230	10	US-09-920-300A-96	Sequence 96, App
C	42	25	1.9	62	10	US-09-983-965-4296	Sequence 4296, App
C	43	25	1.9	77	10	US-09-764-846-282	Sequence 282, App
C	44	25	1.9	95	10	US-09-960-352-8975	Sequence 8975, App
C	45	25	1.9	111	9	US-09-925-299-635	Sequence 635, App

ALIGNMENTS

RESULT 1
US-09-809-545A-1
Sequence 1, Application US/09809545A
Patent No. US2002010804A1
GENERAL INFORMATION:
APPLICANT: Stanton, Lawrence W.
APPLICANT: White, R. Tyler
TITLE OF INVENTION: SECRETED FACTORS
FILE REFERENCE: SCIOS.017A
CURRENT APPLICATION NUMBER: US/09/809,545A
CURRENT FILING DATE: 2001-03-14
NUMBER OF SEQ ID NOS: 84
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1340
TYPE: DNA
ORGANISM: Rattus norvegicus
US-09-809-545A-1

Query Match 100.0%; Score 1340; DB 10; Length 1340;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1340; Conservatve 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GGGCGCCGCTTGAACACATAGCTGCTAGCGCACTTGCCTCACCCAGA	60
DB	1	GGGCGCCGCTTGAACACATAGCTGCTAGCGCACTTGCCTCACCCAGA	60
QY	61	ATGGCATCTCGAGAAATACAGGCGCCCTCATCCCATCCGGCCAGACACCGGCGAG	120
DB	61	ATGGCATCTCGAGAAATACAGGCGCCCTCATCCCATCCGGCCAGACACCGGCGAG	120
QY	121	CCGACTGCTCCCGCCACATTAATCTGATCTCTACAGACGACGCTCGAGACAG	180
DB	121	CCGACTGCTCCCGCCACATTAATCTGATCTCTACAGACGACGCTCGAGACAG	180
QY	181	CGCTTATACAGGCGACAGCGCTCTCCGGACCGCCACAGACAGATGATGCCGCCCG	240
DB	181	CGCTTATACAGGCGACAGCGCTCTCCGGACCGCCACAGACAGATGATGCCGCCCG	240
QY	241	ACGACGCGCGACCGCCACAGACAACTTGTGAATAACAGAAACAAAGTCCCGCCAG	300
DB	241	ACGACGCGCGACCGCCACAGACAACTTGTGAATAACAGAAACAAAGTCCCGCCAG	300

THIS PAGE BLANK (USPTO)

```

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 371..1708
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 371..1705
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-327-035-3

Query Match      1.9%; Score 26; DB 4; Length 2604;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1290 AACGAAAAAAAAAAAAAAAAAAAA 1315
    ||||||||||||||||||||
Db 2538 AACGAAAAAAAAAAAAAAAAAAAA 2563

```

Search completed: March 16, 2003, 02:32:00
 Job time : 99.6285 secs

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/838,399
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Greenfield Ph.D., Michael S.
REGISTRATION NUMBER: 37,147
REFERENCE/DOCKET NUMBER: 96,149/WH 405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)715-1000
TELEFAX: (312)715-1234
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2604 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 371..1708
NAME/KEY: mat_peptide
LOCATION: 371..1705
US-08-838-399-3

Query Match 1.9%; Score 26; DB 2; Length 2604;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1290 AACGAAAAAAAAAAAAAAAAAAAA 1315
|||
Db 2538 AACGAAAAAAAAAAAAAAAAAAAA 2563

RESULT 14
US-09-235-839-3
Sequence 3, Application US/09235839
Patent No. 6207799
GENERAL INFORMATION:
APPLICANT: Hu Ph.D., Yinghe
APPLICANT: McCalieb Ph.D., Michael L.
APPLICANT: Bloomquist Ph.D., Brian T.
APPLICANT: Flores-Riveros Ph.D., Jaime R.
APPLICANT: Cornfield Ph.D., Linda J.
TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
TITLE OF INVENTION: Sequences
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESSES:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive, 32nd Floor
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/235,839
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/630,118
FILING DATE: April 8, 1996
ATTORNEY/AGENT INFORMATION:

NAME: Greenfield Ph.D., Michael S.
REGISTRATION NUMBER: 37,142
REFERENCE/DOCKET NUMBER: 96,149-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)913-0001
TELEFAX: (312)913-0002
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2604 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 371..1705
US-09-235-839-3

Query Match 1.9%; Score 26; DB 4; Length 2604;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1290 AACGAAAAAAAAAAAAAAAAAAAA 1315
|||
Db 2538 AACGAAAAAAAAAAAAAAAAAAAA 2563

RESULT 15
US-09-327-035-3
Sequence 3, Application US/09327035
Patent No. 6368824
GENERAL INFORMATION:
APPLICANT: Hu Ph.D., Yinghe
APPLICANT: McCalieb Ph.D., Michael L.
APPLICANT: Bloomquist Ph.D., Brian T.
APPLICANT: Flores-Riveros Ph.D., Jaime R.
APPLICANT: Cornfield Ph.D., Linda J.
TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
TITLE OF INVENTION: Sequences
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESSES:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/327,035
FILING DATE: 07-Jun-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/838,399
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Greenfield Ph.D., Michael S.
REGISTRATION NUMBER: 37,147
REFERENCE/DOCKET NUMBER: 96,149/WH 405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)715-1000
TELEFAX: (312)715-1234
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2604 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

US-09-327-035-1
Sequence 1, Application US/09327035
Patent No. 6368824
GENERAL INFORMATION:
APPLICANT: Hu Ph.D., Yinghe
McCaleb Ph.D., Michael L.
Bloomquist Ph.D., Brian T.
Flores-Riveros Ph.D., Jaime R.
Cornfield Ph.D., Linda J.
TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
Sequences
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/327,035
FILING DATE: 07-Jun-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/838,399
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Greenfield Ph.D., Michael S.
REGISTRATION NUMBER: 37,147
REFERENCE/DOCKET NUMBER: 96,149/WH 405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)715-1000
TELEFAX: (312)715-1234
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2481 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 248..1585
NAME/KEY: mat_peptide
LOCATION: 248..1582
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-327-035-1
Query Match 1.9%; Score 26; DB 4; Length 2481;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1290 AACGAAAAAAAAAAAAAAAAAAAA 1315
DB 2415 AACGAAAAAAAAAAAAAAAAAAAA 2440

RESULT 12
US-08-630-118A-3
Sequence 3, Application US/08630118A
Patent No. 5919901
GENERAL INFORMATION:
APPLICANT: Hu Ph.D., Yinghe
McCaleb Ph.D., Michael L.
Bloomquist Ph.D., Brian T.

APPLICANT: Flores-Riveros Ph.D., Jaime R.
Cornfield Ph.D., Linda J.
TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
Sequences
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive, 32nd Floor
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,118A
FILING DATE: April 8, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Greenfield Ph.D., Michael S.
REGISTRATION NUMBER: 37,142
REFERENCE/DOCKET NUMBER: 96,149/WH 405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)715-1234
TELEFAX: (312)715-1000
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2604 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 371..1705
US-08-630-118A-3
Query Match 1.9%; Score 26; DB 2; Length 2604;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1290 AACGAAAAAAAAAAAAAAAAAAAA 1315
DB 2538 AACGAAAAAAAAAAAAAAAAAAAA 2563

RESULT 13
US-08-838-399-3
Sequence 3, Application US/08838399
Patent No. 5965392
GENERAL INFORMATION:
APPLICANT: Hu Ph.D., Yinghe
McCaleb Ph.D., Michael L.
Bloomquist Ph.D., Brian T.
Flores-Riveros Ph.D., Jaime R.
Cornfield Ph.D., Linda J.
TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
Sequences
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

LENGTH: 2481 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 248..1582
US-08-630-118A-1

Query Match 1.9%; Score 26; DB 2; Length 2481;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1290 AACGAAAAAAAAAAAAAAAAAAAA 1315
DB 2415 AACGAAAAAAAAAAAAAAAAAAAA 2440

RESULT 9

US-08-838-399-1
Sequence 1, Application US/08838399
Patent No. 5965392

GENERAL INFORMATION:

APPLICANT: Hu Ph.D., Yinghe
APPLICANT: McCaleb Ph.D., Michael L.
APPLICANT: Bloomquist Ph.D., Brian T.
APPLICANT: Flores-Riveros Ph.D., Jaime R.
APPLICANT: Cornfield Ph.D., Linda J.
TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/838,399
FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Greenfield Ph.D., Michael S.
REGISTRATION NUMBER: 37,147
REFERENCE/DOCKET NUMBER: 96,149/WH 405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)715-1000
TELEFAX: (312)715-1234
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2481 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO

FEATURE:

NAME/KEY: CDS
LOCATION: 248..1585
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 248..1582
US-08-838-399-1

Query Match 1.9%; Score 26; DB 2; Length 2481;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1290 AACGAAAAAAAAAAAAAAAAAAAA 1315
DB 2415 AACGAAAAAAAAAAAAAAAAAAAA 2440

RESULT 10

US-09-235-839-1
Sequence 1, Application US/09235839
Patent No. 6207799

GENERAL INFORMATION:

APPLICANT: Hu Ph.D., Yinghe
APPLICANT: McCaleb Ph.D., Michael L.
APPLICANT: Bloomquist Ph.D., Brian T.
APPLICANT: Flores-Riveros Ph.D., Jaime R.
APPLICANT: Cornfield Ph.D., Linda J.
TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive, 32nd Floor
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/235,839
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/630,118
FILING DATE: April 8, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Greenfield Ph.D., Michael S.
REGISTRATION NUMBER: 37,142
REFERENCE/DOCKET NUMBER: 96,149-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)913-0001
TELEFAX: (312)913-0002
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2481 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO

FEATURE:

NAME/KEY: CDS
LOCATION: 248..1582
US-09-235-839-1

Query Match 1.9%; Score 26; DB 4; Length 2481;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1290 AACGAAAAAAAAAAAAAAAAAAAA 1315
DB 2415 AACGAAAAAAAAAAAAAAAAAAAA 2440

RESULT 11

RESULT 6
US-08-911-020-1
Sequence 1, Application US/08911020
Patent No. 5854047
GENERAL INFORMATION:
APPLICANT: Buchert, Johanna
APPLICANT: Silka-aho, Martti
APPLICANT: Viikari, Liisa
APPLICANT: Penttila, Merja
APPLICANT: Saloheimo, Anu
APPLICANT: Marjatta, Ranna
TITLE OF INVENTION: Mannanase enzymes, genes coding for them,
TITLE OF INVENTION: methods for isolating the genes, and methods for bleaching
TITLE OF INVENTION: lignocellulosic pulps
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolaesch and Birch
STREET: PO Box 747
CITY: Falls Church
STATE: VA
COUNTRY: US
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/911,020
FILING DATE: 13-Aug-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/341,568
FILING DATE: 22-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr, Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 365-262P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 536 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Trichoderma reesei
STRAIN: QM9414
US-08-911-020-1

Query Match 1.9%; Score 26; DB 2; Length 536;
Best Local Similarity 100.0%; Pred. No. 0.037;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1290 AACGAAAAAAAAAAAAAAAAAAAA 1315
DB 460 AACGAAAAAAAAAAAAAAAAAAAA 485

RESULT 7
US-09-328-111-361/C
Sequence 361, Application US/09328111
Patent No. 6262333
GENERAL INFORMATION:
APPLICANT: Endegge, Wilson O.
APPLICANT: Steinmann, Kathleen E.
APPLICANT: Aetle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Buehneil, Steven E.

APPLICANT: Carroll III, Eddie
APPLICANT: Catino, Theodore J.
APPLICANT: Dertli, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
TITLE OF INVENTION: PRODUCTS
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328,111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 361
LENGTH: 631
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)..(631)
OTHER INFORMATION: n = A,T,C or G
US-09-328-111-361

Query Match 1.9%; Score 26; DB 4; Length 631;
Best Local Similarity 100.0%; Pred. No. 0.037;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1290 AACGAAAAAAAAAAAAAAAAAAAA 1315
DB 44 AACGAAAAAAAAAAAAAAAAAAAA 19

RESULT 8
US-08-630-118A-1
Sequence 1, Application US/08630118A
Patent No. 5919901
GENERAL INFORMATION:
APPLICANT: Hu Ph.D., Yinghe
APPLICANT: McCalab Ph.D., Michael L.
APPLICANT: Bloomquist Ph.D., Brian T.
APPLICANT: Flores-Riveros Ph.D., Jaime R.
APPLICANT: Cornfield Ph.D., Linda J.
TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
TITLE OF INVENTION: Sequences
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive, 32nd Floor
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,118A
FILING DATE: April 8, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Greenfield Ph.D., Michael S.
REGISTRATION NUMBER: 37,142
REFERENCE/DOCKET NUMBER: 96,149/WH 405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)715-1000
TELEFAX: (312)715-1234
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

;; CURRENT FILING DATE: 1998-10-21
;; NUMBER OF SEQ ID NOS: 9
;; SOFTWARE: PERL Program
;; SEQ ID NO 5
;; LENGTH: 1506
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: 1250374
US-09-176-657-5

Query Match 2.2%; Score 29; DB 3; Length 1506;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 484 ACCGTGTAGAGGCGCGTAAATCGAGGT 512
Db 717 ACCGTGTAGAGGCGCGTAAATCGAGGT 745

RESULT 3

US-09-306-290-6/c
;; Sequence 6, Application US/09306290
;; Patent No. 6221635
;; GENERAL INFORMATION:
;; APPLICANT: Rovera, Giovanni
;; APPLICANT: Mukhopadhyay, Sunil
;; TITLE OF INVENTION: METHODS FOR SOLID-PHASE AMPLIFICATION OF DNA TEMPLATE
;; FILE REFERENCE: 09924-10
;; CURRENT APPLICATION NUMBER: US/09/306,290
;; CURRENT FILING DATE: 1999-05-06
;; NUMBER OF SEQ ID NOS: 43
;; SOFTWARE: Patentin Ver. 2.0
;; SEQ ID NO 6
;; LENGTH: 40
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: Primer RGP
US-09-306-290-6

Query Match 1.9%; Score 26; DB 4; Length 40;
Best Local Similarity 100.0%; Pred. No. 0.044;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1288 ACAACGAAAAAAAAAAAAAAAAAAAA 1313

Db 26 ACAACGAAAAAAAAAAAAAAAAAAAA 1

RESULT 4

US-09-105-542A-12
;; Sequence 12, Application US/09105542A
;; Patent No. 6323329
;; GENERAL INFORMATION:
;; APPLICANT: Bullerdiek, Jörn
;; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES OF GENES OF THE
;; TITLE OF INVENTION: HIGH MOBILITY GROUP PROTEINS AND USES THEREFORE
;; FILE REFERENCE: BOEHM4.001C1P
;; CURRENT APPLICATION NUMBER: US/09/105,542A
;; CURRENT FILING DATE: 1998-06-26
;; PRIOR APPLICATION NUMBER: PCT/DE96/02494
;; PRIOR FILING DATE: 1996-12-20
;; NUMBER OF SEQ ID NOS: 23
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 12
;; LENGTH: 528
;; TYPE: DNA
;; ORGANISM: Homo Sapien
;; FEATURE:
;; NAME/KEY: misc_feature

;; LOCATION: (1)...(528)
;; OTHER INFORMATION: n = A,T,C or G
US-09-105-542A-12

Query Match 1.9%; Score 26; DB 4; Length 528;
Best Local Similarity 100.0%; Pred. No. 0.037;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1290 AACGAAAAAAAAAAAAAAAAAAAA 1315
Db 472 AACGAAAAAAAAAAAAAAAAAAAA 497

RESULT 5

US-08-341-568-1
;; Sequence 1, Application US/08341568
;; Patent No. 5661021
;; GENERAL INFORMATION:
;; APPLICANT: Buchert, Johanna
;; APPLICANT: Silka-aho, Matti
;; APPLICANT: Viikari, Liisa
;; APPLICANT: Penttilä, Merja
;; APPLICANT: Saloheimo, Anna
;; APPLICANT: Marjatta, Ranna
;; TITLE OF INVENTION: Mannanase enzymes, genes coding for them,
;; TITLE OF INVENTION: methods for isolating the genes, and methods for bleaching lign
;; TITLE OF INVENTION: pulps
;; NUMBER OF SEQUENCES: 7
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Birch, Stewart, Kolaesch and Birch
;; STREET: PO Box 747
;; CITY: Falls Church
;; STATE: VA
;; COUNTRY: US
;; ZIP: 22040-0747

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/341,568
FILING DATE: 22-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr, Gerald M
REGISTRATION NUMBER: 28,977
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 536 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Trichoderma reesei
STRAIN: QM9414
US-08-341-568-1

Query Match 1.9%; Score 26; DB 1; Length 536;
Best Local Similarity 100.0%; Pred. No. 0.037;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1290 AACGAAAAAAAAAAAAAAAAAAAA 1315
Db 460 AACGAAAAAAAAAAAAAAAAAAAA 485

GenCore version 5.1.4.J5_4578
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 16:12:15 ; Search time 75.6265 Seconds
(without alignments)
5433.755 Million cell updates/sec

Title: US-09-809-545A-1

Perfect score: 1340

Sequence: 1 ggcggccggccctgcacacat.....ggccgaagggtcctaga 1340

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 15338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patente.NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by change to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	70	5.2	2372	4	US-09-145-391-1
2	29	2.2	1506	3	US-09-176-657-5
3	26	1.9	40	4	US-09-306-290-6
4	26	1.9	528	4	US-09-105-542A-12
5	26	1.9	536	1	US-08-341-568-1
6	26	1.9	536	2	US-08-911-020-1
7	26	1.9	631	4	US-09-328-111-61
8	26	1.9	2481	2	US-08-630-118A-1
9	26	1.9	2481	2	US-08-838-399-1
10	26	1.9	2481	4	US-09-235-839-1
11	26	1.9	2481	4	US-09-327-035-1
12	26	1.9	2604	2	US-08-630-118A-3
13	26	1.9	2604	2	US-08-838-399-3
14	26	1.9	2604	4	US-09-235-839-3
15	26	1.9	2604	4	US-09-327-035-3
16	26	1.9	2610	4	US-09-545-814-1
17	26	1.9	2610	4	US-09-545-814-3
18	26	1.9	2738	2	US-08-795-868-17
19	26	1.9	2738	4	US-09-303-069-17
20	26	1.9	2738	4	US-09-134-250-17
21	26	1.9	3487	4	US-09-303-069-24
22	26	1.9	7291	4	US-08-913-942-3
23	25	1.9	291	4	US-09-329-796-1
24	25	1.9	572	4	US-09-342-653-5
25	25	1.9	599	4	US-09-342-647-19
26	25	1.9	608	4	US-09-385-982-183
27	25	1.9	903	5	PCT-US95-06406A-21

28	25	1.9	1074	4	US-09-357-251-23	Sequence 23, Appl
29	25	1.9	1132	3	US-08-894-731-3	Sequence 3, Appl
30	25	1.9	1375	4	US-09-372-422A-37	Sequence 37, Appl
31	25	1.9	1543	2	US-08-837-593-1	Sequence 1, Appl
32	25	1.9	1584	1	US-08-361-920-24	Sequence 24, Appl
33	25	1.9	1584	1	US-08-479-939-24	Sequence 24, Appl
34	25	1.9	1584	1	US-08-483-432-24	Sequence 24, Appl
35	25	1.9	2071	4	US-09-816-088-1	Sequence 1, Appl
36	25	1.9	2885	2	US-08-791-849A-15	Sequence 15, Appl
37	25	1.9	2901	4	US-08-368-776A-4	Sequence 4, Appl
38	25	1.9	2901	5	PCT-US96-00419-4	Sequence 4, Appl
39	25	1.9	3180	1	US-08-480-662-1	Sequence 1, Appl
40	25	1.9	3180	3	US-08-918-190-1	Sequence 1, Appl
41	25	1.9	3180	4	US-09-234-232-1	Sequence 1, Appl
42	25	1.9	3180	5	PCT-US96-09927-1	Sequence 1, Appl
43	25	1.9	3238	4	US-08-123-934A-5	Sequence 5, Appl
44	25	1.9	3238	5	PCT-US94-10080-5	Sequence 5, Appl
45	25	1.9	3410	4	US-09-020-956-110	Sequence 110, Appl

ALIGNMENTS

RESULT 1
US-09-145-391-1
Sequence 1, Application US/09145391
Patent No. 6194171
GENERAL INFORMATION:
APPLICANT: Pulist, Stefan M.
TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
TITLE OF INVENTION: Products Related Thereto and Methods of Using Same
FILE REFERENCE: CE 3093
CURRENT APPLICATION NUMBER: US/09/145,391
CURRENT FILING DATE: 1998-09-01
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 2372
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (987) ..(1979)
US-09-145-391-1

Query Match 5.2%; Score 70; DB 4; Length 2372;
Best Local Similarity 100.0%; Pred.No. 6.4e-21;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1045 GATCCAAAGACTAGAGCCATGCTGATGATGAGTGGTCTGTTCTTTCTTCAATTCAGAGCT 1104
Db 2133 GATCCAAAGACTAGAGCCATGCTGATGATGAGTGGTCTGTTCTTTCTTCAATTCAGAGCT 2192

Qy 1105 AGTATATACC 1114
Db 2193 AGTATATACC 2202

RESULT 2
US-09-176-657-5
Sequence 5, Application US/09176657
Patent No. 6020164
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Tang, Y. Tom
APPLICANT: Corley, Neil C.
APPLICANT: Guejler, Karl J.
APPLICANT: Lu, Aina
APPLICANT: Baughn, Mariah R.
TITLE OF INVENTION: HUMAN RNA BINDING PROTEINS
FILE REFERENCE: PF-0611 US
CURRENT APPLICATION NUMBER: US/09/176,657

THIS PAGE BLANK (USPTO)

BASE COUNT	108 a	104 c	91 g	101 t	4 others
ORIGIN					
Query Match	6.0%; Score 80; DB 12; Length 1408;				
Best Local Similarity	100.0%; Pred. No. 0;				
Matches	80; Conservative	0; Mismatches	0; Indels	0; Gaps	0; 0;
Qy	556	AACCCCTACACCAATGCTGTAATTAATTCAGTTGTGGGCGCGTCAAGCCCGAC	615		
Db	232	AACCCCTACACCAATGCTGTAATTAATTCAGTTGTGGGCGCGTCAAGCCCGAC	291		
Qy	616	TTCTATGCAGGACGCGTCT	635		
Db	292	TTCTATGCAGGACGCGTCT	311		
RESULT 15					
LOCUS	228900	228 bp	mRNA	linear	EST 14-DEC-1993
DEFINITION	HSHB7F031 STRATAGENE Human skeletal muscle cDNA library, cat.				
ACCESSION	228900	#936215. Homo sapiens cDNA clone B7F03, mRNA sequence.			
VERSION	228900.1	GI:434570	EST.		
KEYWORDS	human.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
AUTHORS	1 (bases 1 to 228) Aufrey,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes,M.D., Duprat,S., Houligate,R., Jumeau,M.N., Lamy,B., Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pletu,G., Pouillot,Y., Sebasteiani-Kabackchis,C. and Tessier,A.				
TITLE	IMAGE: molecular integration of the analysis of the human genome and its expression				
JOURNAL	C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)				
MEDLINE	95277534				
COMMENT	Contact: Genethon Genethon Genethon Centre de recherche sur le Genome Humain 1, rue de l'Internationale, Bp60 91002 EVRY Cedex, FRANCE Tel: 33169472800 Fax: 33160778698 Email: genethon@genethon.fr strand(-), single read.				
FEATURES	Location/Qualifiers				
Source	1..228				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="B7F03"				
	/clone_lib="STRATAGENE Human skeletal muscle cDNA library,"				
	cat. #936215."				
	/sex="Female"				
	/tissue_type="skeletal muscle"				
	/dev_stage="19 years"				
	/note="Organ: leg muscle; Vector: lambda ZAPII; Tissue from female, 19 years old, normal leg muscle. Cloning vector is lambda ZAPII, in vivo excision from lambda ZAPIII				

```

BASE COUNT      64 a      51 c      48 g      64 t      1 others
ORIGIN
Query Match      5.2%; Score 70; DB 14; Length 228;
Best Local Similarity 100.0%; Pired. No. 0;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1045 GATGCCAAGACTAGAGAGCCATGTCGATGATGAGGCTCGTCTCTTTCTTCATTGCAGGCT 1104
DB 201 GATGCCAAGACTAGAGAGCCATGTCGATGATGAGGCTCTGCTTTCTTCATTGCAGGCT 142
QY 1105 AGTATATACC 1114
DB 141 AGTATATACC 132
Search completed: March 16, 2003, 02:29:40
Job time : 1895.21 secs

```

```
Search completed: March 16, 2003, 02:29:40
Job time : 1895.21 secs
```

```

/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-Y0-acb-d-03-0-UI"
/clone_lib="UI-R-Y0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-Y0
library is a subtracted library derived from an
individually-tagged normalized whole-eye (minus the lens)
library. The driver for the subtraction consisted of a
pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-B0,
UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of
3-5 nucleotides present between the Not I site and the
oligo-dt track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (UI-R-Y0) was constructed as follows: PCR
amplified cDNA inserts from previous library clones from
which 3' ESTs had been derived were used as a driver in a
hybridization with the normalized whole-eye library in
the form of single-stranded circles. The remaining
single-stranded circles (subtracted library) was purified
by hydroxypapate column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the UI-R-Y0
library. This procedure has been previously described
(Bonardo, Lennon and Soares, Genome Research 6: 791-806,
1996)
TAG_LIB=UI-R-Y0
TAG_TISSUE=EYE
TAG_SEQ=CATTTG"
BASE COUNT      103 a      81 g      64 t
ORIGIN

```

```

Query Match      10.4%; Score 140; DB 9; Length 395;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 200 CCGTCTCGGACCGCCACAGACAGATGATCGCCCGCCGACGAGCCGACCCAGCA 259
      |||||||
DB 205 CCGTCTCGGACCGCCACAGACAGATGATCGCCCGCCGACGAGCCGACCCAGCA 264
      |||||||
QY 260 CACAACCTTGTGAAAAACAGAAAAACAAGTCCAGCCCAAGGCGTGCATGTGTCAACA 319
      |||||||
DB 265 CACAACCTTGTGAAAAACAGAAAAACAAGTCCAGCCCAAGGCGTGCATGTGTCAACA 324
      |||||||
QY 320 TCCCTCTCGGTTCCGGAT 339
      |||||||
DB 325 TCCCTCTCGGTTCCGGAT 344
      |||||||

```

```

RESULT 13
BI988957 600 bp mRNA linear EST 20-DEC-2001
LOCUS 4023-45 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
DEFINITION mRNA sequence.
ACCESSION BI988957
VERSION BI988957.1 GI:17959947
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 600)
AUTHORS Mu,X., Zhao,S., Pershad,R., Heieh,T.-F., Scarpa,A., Wang,S.W.,
TITLE White,R.A., Beremand,P.D., Thomas,T.B., Gan,L. and Klein,W.H.
JOURNAL Gene expression in the developing mouse retina by EST sequencing
MEDLINE and microarray analysis
COMMENT Nucleic Acids Res. 29 (24), 4983-4993 (2001)
CONTACT: Klein WH
DEPARTMENT: Biochemistry and Molecular Biology
UNIVERSITY: University of Texas M.D. Anderson Cancer Center

```

```

Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
Tel: 713 792 3646
Fax: 713 790 0329.
Location/Qualifiers
1..600
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="Mouse E14.5 retina lambda ZAP II Library"
/tissue_type="neural retina"
/dev_stage="embryonic day 14.5 post-fertilization"
/notes="Vector: pAMP10 (Gibco); Cloned unidirectionally.
Primer: Oligo dT. RNA Isolation: cytoplasmic RNA preps
(Mammatis); Cloning technique: cDNA Cloning (Clontekmp,
Life Technologies); Average insert size: 1.8 Kb;
Insertion site: TAGCTCCACTGAAATCTGAGTG--->. Other
information regarding entire library may be found at
http://pga.sommed.edu/data/libraries/microarray_cdna_library.htm."

```

```

BASE COUNT      171 a      124 c      122 g      183 t
ORIGIN

```

```

Query Match      7.2%; Score 97; DB 13; Length 600;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 147; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1023 GAATGCTTTTGGCCCTTGACCGATGCAAGACTAGAGCCATGCTGATGATGCGTCT 1082
      |||||||
DB 450 GAATGCTTTTGGCCCTTGACCGATGCAAGACTAGAGCCATGCTGATGATGCGTCT 509
      |||||||
QY 1083 CGTCTCTTCTTCATTGACGCTAGATATATACCAAGGGGATCAACCGTTTGTCCATA 1142
      |||||||
DB 510 CGTCTCTTCTTCATTGACGCTAGATATATACCAAGGGGATCAACCGTTTGTCCATA 569
      |||||||
QY 1143 TTAATGATATAAACCATTAACAAACA 1170
      |||||||
DB 570 TTAATGATATAAACCATTAACAAACA 597
      |||||||

```

```

RESULT 14
BG795922 408 bp mRNA linear EST 16-MAY-2001
LOCUS UTSW_SM385 UTSW Adult Mouse Skeletal Muscle Library Mus musculus
DEFINITION cDNA clone UTSW_SM385, mRNA sequence.
ACCESSION BG795922
VERSION BG795922.1 GI:14131492
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 408)
AUTHORS Williams,R.S. and Shohet,R.V., Persenlidis,A., Garner,H.R.,
TITLE UT Southwestern Medical Center, Adult Mouse Skeletal Muscle cDNA
Library
JOURNAL Unpublished (2001)
COMMENT Contact: Schageman JJ
Shohet/Garner Labs
University of Texas Southwestern Medical Center
6000 Harry Hines Blvd., NA2.226, Dallas, TX 75390, USA
Tel: 214 648 1674
Email: Jeff.Schageman@UTSouthwestern.edu
CDNA library constructed by UTSW as a component of the program for
Genomic Applications (PGA) and the Reynolds Heart Disease
Prevention grants for use in cDNA microarray experiments. Sequence
Quality: Sequence ends were trimmed based on percentage of ambigu-
us base calls or 'N's in windowed segments. Sequencing: First-pass
sequencing; ABI Prism 377 sequencer and analysis software.
Seq primer: M13/pUC Reverse.
Location/Qualifiers
1..408
/organism="Mus musculus"
/db_xref="taxon:10090"

```

ORIGIN

Query Match 11.7%; Score 157; DB 9; Length 361;
 Best Local Similarity 99.5%; Pred. No. 0;
 Matches 207; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 963 TGCTGCGGACCCCTTACACACACACCTTGTCCAGCCCTTACGGCGTGTGTCAT 1022
 DB 358 TGCTGCGGACCCCTTACACACACACCTTGTCCAGCCCTTACGGCGTGTGTCAT 299
 QY 1023 GAATGCTTTTGGCCCTTGACCGATGCGCAAGATGAGGACCATGCTGATGTGGCTCT 1082
 DB 298 GAATGCTTTTGGCCCTTGACCGATGCGCAAGATGAGGACCATGCTGATGTGGCTCT 239
 QY 1083 CGTTCTTTCTTTCATGAGCGCTAGTATATACCAAGGGGATACCAACCGTTTGTCTCCATA 1142
 DB 238 CGTTCTTTCTTTCATGAGCGCTAGTATATACCAAGGGGATACCAACCGTTTGTCTCCATA 179
 QY 1143 TTAATGATTAACCATTAACCAACCA 1170
 DB 178 TTAATGATTAACCATTAACCAACCA 151

RESULT 11

LOCUS BB750925 425 bp mRNA linear EST 16-OCT-2001
 DEFINITION BB750925 RIKEN full-length enriched, pooled tissues, cerebellum,
 etc. Mus musculus cDNA clone G130203B11 3', mRNA sequence.
 BB750925
 BB750925.1 GI:16155161

ACCESSION

KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE

Akimura, T., Arikawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Warahashi, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)

JOURNAL

COMMENT Unpublished (2001)
 Contact: Yoshitake Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suenro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-resgsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)

Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

FEATURES

Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.
 e mouse tissues.
 Location/Qualifiers
 1..425
 /organism="Mus musculus"
 /strain="G57BL/6J"
 /db_xref="taxon:10090"
 /clone="G130203B11"
 /clone_1b="RIKEN full-length enriched, pooled tissues,
 cerebellum, etc."
 /note="pooled tissues; (tissue_type=cerebellum,
 dev_stage=16 days neonate, sex=mixed),
 (tissue_type=cerebellum, dev_stage=0 day neonate,
 sex=mixed), (tissue_type=hippocampus, dev_stage=adult,
 sex=male), (tissue_type=whole body, dev_stage=9 days
 embryo, sex=mixed) (tissue_type=lung, dev_stage=13 days
 embryo, sex=mixed)"

BASE COUNT

128 a 109 c 87 g 101 t

Query Match 10.9%; Score 146; DB 10; Length 425;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 945 CAGTTACGACGAGATTATGCTCCGACCCCTTACACACACACTTGTCCAGCCCCC 1004
 DB 73 CAGTTACGACGAGATTATGCTCCGACCCCTTACACACACACTTGTCCAGCCCCC 132
 QY 1005 CTACGGCGTTGGTGCATGATGCTTTGGCCCTTACCGAGCCAAAGCTGAGAGCCA 1064
 DB 133 CTACGGCGTTGGTGCATGATGCTTTGGCCCTTACCGAGCCAAAGCTGAGAGCCA 192
 QY 1065 TGCTGATGATGCTGCTCTCTCTTT 1090
 DB 193 TGCTGATGATGCTGCTCTCTCTTT 218

RESULT 12

LOCUS A1717457 395 bp mRNA linear EST 10-JUN-1999
 A1717457
 DEFINITION UI-R-Y0-acd-d-03-0-UI-R1 UI-R-Y0 Rattus norvegicus cDNA clone
 UI-R-Y0-acd-d-03-0-UI 3', mRNA sequence.
 A1717457
 A1717457.1 GI:5034713

ACCESSION

KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 395)
 Ronaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477

JOURNAL

MEDLINE
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu

Oligo-dt track not found. Not a site shown in beginning of sequence
 is likely internal to the message. cDNA Library Preparation: M.B.
 Soares Lab Clone distribution: Clones will be available through
 Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA-No.

FEATURES

source Location/Qualifiers
 1..395
 /organism="Rattus norvegicus"

Qy	1143	TTAAATATATAAACCATTTAAACAAACAGCAAAAACAAACCAAAATATCAAC	1202
Db	164	TTAAATATATAAACCATTTAAACAAACAAACAAACCAAAATATCAAC	105
Qy	1203	CTTCCAAATGTGGGAGAGAGAAAGCTTCCGAGGCCGAGTGTGGACATCGCATAG	1262
Db	104	CTTCCAAATGTGGGAGAGAGAAAGCTTCCGAGGCCGAGTGTGGACATCGCATAG	45
Qy	1263	GACATCACTTTAGCAACTCAAGAAACAC	1292
Db	44	GACATCACTTTAGCAACTCAAGAAACAC	15

RESULT	9
BFS49922/c	
LOCUS	BFE49922
DEFINITION	UI-R-E0-DV-C-04-0-UI.#1 UI-R-E0 Rattus norvegicus cdna clone
ACCESSION	UI-R-E0-DV-C-04-0-UI.5, mRNA sequence.
VERSION	BFE49922
KEYWORDS	BFE49922.1 GI:11659652
SOURCE	EST.
ORGANISM	Norway rat. Rattus norvegicus

REFERENCE 1 (bases 1 to 533)
AUTHORS Ronaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
JOURNAL discovery
MEDLINE Genome Res. 6 (9), 791-806 (1996)
COMMENT 97044477
Contact: Soares, MB

University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel.: 319 335 8250
Fax: 319 335 9565
Email: msatres@blue.weeg.uiowa.edu
CDNA Library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.rtggen.com)
This clone is also available through the I.M.A.G.E. Consortium at
LNLN (info@image.lln.gov). IMAGE ID=1777166
Seq primer: M3 forward.

FEATURES	source
Location/Qualifiers	1..533
/organism="Rattus norvegicus"	
/strain="Sprague-Dawley"	
/db_xref="taxon:10116"	
/clone="UI-R-E0-bv-C-04-0-UI"	
/clone_1b="UI-R-E0"	
/dev_stage="embryonic"	
/lab_host="DH10B (Life Technologies)"	
/note="Vector: pTR73D-pac (Pharmacia) with a modified polylinker; Site_1: NotI; Site_2: EcoRI; This library consists of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."	
	134 a 125 c 164 g 109 t 1 others

Query Match	11.9%	Score 160;	DB 12;	Length 533;
Best Local Similarity	99.5%	Pred. No. 0;		
Matches 210; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0

Qy 904 TACGCCACGCCACCCCTCTGCTCTGCTCTACAGTACAGTTACGAGAGATTAT 963

Db 268 TACGCCACGCCACCCCTCTGCTCTGCTCTCTACAGTACAGTTACGAGAGATTAT 209

Qy 964 GCTGCCAGCCCTACACACACACACTTGTCTCAACCCCCACCTACGGCGTTGTGTGCATG 1023

	108	148	89
D _b	208	GCTCCGACCCCTACACACACTTGCTCAGCCCCCATCGCGCGTGGGCATG	
O _y	1024	AATGCTTTGGGCGCCCTGACCGAATGCCAAGACTAGAGCAGTGTGATGTGGGCTC	108
D _b	148	AATGCTTTGGGCGCCCTGACCGAATGCCAAGACTAGAGCAGTGTGATGTGGGCTC	89
O _y	1084	GTTCTTCTTCATGACGAGGTAGATATACC	1114
D _b	88	GTTCTTCTTCATGACGAGGTAGATATACC	58

RESULT 10	LOCUS	DEFINITION	VERSION	KEYWORDS	SOURCE	ORGANISM
AI835011/c	361 bp	mus musculus cDNA clone	AI835011	UI-M-AM1-afw-h-01-0-UI s1 NIH BMAP_MAM_Nus	house mouse.	Chordata: Craniata: Vertebrata: Euteleostomi: Euteleostei: Muraena: Chordata: Vertebrata: Euteleostomi:
		mus musculus cDNA clone	AI835011	UI-M-AM1-afw-h-01-0-UI 3', mRNA sequence.	house mouse.	
			AI835011.1	GI:5469224	house mouse.	
			EST.		house mouse.	
					mus musculus	

REFERENCE	1 (bases 1 to 361)
AUTHORS	Ronald M.F., Lennon G. and Soares M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	97044477
COMMENT	Contact: Chin, H

JOURNAL
MEDLINE
COMMENT

Genome Res. 6 (9), 791-806(1996)
97044477
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: MEST@mail.nih.gov

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized amygdala library cDNA Library Preparation: M.B. Soares lab Clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined. The following repetitive elements were found in this cDNA sequence: 122-168, >(CMAA)n\$simple_repeat
Seq primer: M13 Forward
POLYA=Yes.

FEATURES	Location/Qualifiers
source	1. .361

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-AM1-afw-h-01-0-UI"
/clone_lib="NIH_BMAP_MAM_N"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pRT13D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NIH BMAP MAM_N library is a normalized library constructed
from mouse amygdala. The tag is a string of 5 nucleotides
present between the Not I site and the Oligo-dT track.
The library was constructed as described by Bonaldo,
Lemon and Soares, Genome Research 6: 791-806, 1996.
Tissue provided by Ms. Annie Novakovich, Zivic-Miller
laboratories.
TAG_LIB=NIH_BMAP_MAM_N
TAG_TISSUE=amygdala
TAG_SEQ=GTGAG

```

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 430)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: MEST@mail.nih.gov
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a donatide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized spinal cord library cDNA library Preparation: M.B. Soares Lab Clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined. The following repetitive elements were found in this cDNA sequence:
116-162, >(CAAA)n\$Simple_repeat
Seq primer: M13 Forward
POLYA=Yes

FEATURES
source Location/Qualifiers
1..430
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="UI-M-BG1-ale-e-06-0-UI"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pTZ19-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; The NIH_BMAP_MSC_N library is a normalized library constructed from mouse spinal cord. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares. Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories.
TAG_LIB=NIH_BMAP_MSC_N
TAG_TISSUE=spinal-cord
TAG_SEQ=TCGAA"

BASE COUNT 97 a 85 c 104 g 144 t
ORIGIN

Query Match 13.1%; Score 175; DB 9; Length 430;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 225; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 945 CAGTACGGAGCGATTATGCTGCGGACCCCTACACACACACACTTGCTCAGCCCCAC 1004
DB 370 CAGTACGGAGCGATTATGCTGCGGACCCCTACACACACACACTTGCTCAGCCCCAC 311
QY 1005 CTACGGCGTGTGCGCATGATGCTTTGGCCCTTGACCGATGCCAAGACTAGAGACCA 1064
DB 310 CTACGGCGTGTGCGCATGATGCTTTGGCCCTTGACCGATGCCAAGACTAGAGACCA 251
QY 1065 TGCTGATGATGCGGTCTGCTTTCTTCTTATGACGAGCTAGATATATACCAAGGGGAGTA 1124
DB 250 TGCTGATGATGCGGTCTGCTTTCTTCTTATGACGAGCTAGATATATACCAAGGGGAGTA 191
QY 1125 CAACCGTTTGCTCCATTAATGATAAAACATTAAACAA 1170
DB 190 CAACCGTTTGCTCCATTAATGATAAAACATTAAACAA 145

RESULT 8
AA818804/c
LOCUS
DEFINITION
AA818804 538 bp mRNA linear EST 03-JUL-1999
UI-R-A0-ar-b-11-0-UI s1 UI-R-A0 Rattus norvegicus cDNA clone
UI-R-A0-ar-b-11-0-UI 3, similar to gi|2885286|gb|AA773715|AA773715
a181c05.r1 Soares NHMPPu SI Homo sapiens cDNA clone 1048424 5',
mRNA sequence.
AA818804
AA818804.1 GI:4227480
EST.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 538)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On Feb 17, 1998 this sequence version replaced gi:2888390.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult Brain library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics This clone is also available through the I.M.A.G.E. Consortium at LNMU (info@image.lnm.gov). IMAGE ID=1767889 The following repetitive elements were found in this cDNA sequence:
105-154, >(CAAA)n\$Simple_repeat
Seq primer: M13 Forward
POLYA=No.

FEATURES
source Location/Qualifiers
1..538
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone_lib="UI-R-A0-ar-b-11-0-UI"
/dev_stage="adult"
/note="Vector: pTZ19-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; This library consists of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."

BASE COUNT 146 a 115 c 102 g 175 t
ORIGIN

Query Match 12.5%; Score 168; DB 9; Length 538;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 268; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1023 GAATGCTTTTGGGCGCTTACCGAGTCCCAAGACTAGAGCCATGCTGATGATGTGGTCT 1082
DB 284 GAATGCTTTTGGGCGCTTACCGAGTCCCAAGACTAGAGCCATGCTGATGATGTGGTCT 225
QY 1083 CGTTCTTTCTCATGACGAGCTAGTATATACCAAGGGGAGTCAACCGTTTGCTCCATA 1142
DB 224 CGTTCTTTCTCATGACGAGCTAGTATATACCAAGGGGAGTCAACCGTTTGCTCCATA 165

VERSION A1843387.1 GI:5477600
 EST.
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 379)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: mst@mail.nih.gov
 The sequence contained an oligo-dt track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dt track served to verify it as a clone from the normalized hippocampus library cDNA library Preparation: M.B. Soares Lab Clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined. The following repetitive elements were found in this cDNA sequence: 117-163, >(CAAA)n\$Simple_repeat
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES
 source
 1..379
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-A01-adz-f-10-0-UI"
 /clone_1db="NIH BMAP_MHI_N"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The NIH BMAP MHI N library is a normalized library constructed from mouse hippocampus. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dt track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories.
 TAG_LIB=NIH BMAP_MHI_N
 TAG_TISSUE=hippocampus
 TAG_SEQ=TTTGA"

BASE COUNT 82 a 75 c 92 g 130 t
 ORIGIN

Query Match 13.7%; Score 183; DB 9; Length 379;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 233; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 937 TACACTGACAGTTCAGAGAGATTATGCTGCGACCCCTACACACACACTTGTCTCA 996
 DB 379 TACACTGACAGTTCAGAGAGATTATGCTGCGACCCCTACACACACACTTGTCTCA 320

OY 997 GCGCCACCTACGCGCTGTGTCATGAATGCTTTTGCGCCCTTGACCGATCCCAAGCT 1056
 DB 319 GCGCCACCTACGCGCTGTGTCATGAATGCTTTTGCGCCCTTGACCGATCCCAAGCT 260

OY 1057 AGGAGCCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1116
 DB 259 AGGAGCCATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 200

OY 1117 GGGGATACAAACCGTTTGTCTCATATTAATGATTAACCATTAACCAACAA 1170
 DB 199 GGGGATACAAACCGTTTGTCTCATATTAATGATTAACCATTAACCAACAA 146

RESULT 6
 LOCUS A1175239/c
 DEFINITION EST218774 Normalized rat muscle, Bento Soares Rattus sp. cDNA clone RMUCER0 3' end, mRNA sequence.
 ACCESSION A1175239
 VERSION A1175239.1 GI:3725877
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 REFERENCE 1 (bases 1 to 266)
 AUTHORS Lee,N.H., Glodok,A., Chandra,I., Mason,T.M., Quackenbush,J., Karlavage,A.R. and Adams,M.D.
 TITLE Rat Genome Project: Generation of a Rat EST (RESt) Catalog & Rat Gene Index
 JOURNAL Unpublished (1998)
 COMMENT Contact: Lee, NH
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@igr.org
 Seq primer: M13-21.

FEATURES
 source
 1..266
 /organism="Rattus sp."
 /db_xref="ATCC (inhost):2032297"
 /db_xref="taxon:10118"
 /clone="RMUCER0"
 /clone_1db="Normalized rat muscle, Bento Soares"
 /note="Organ: muscle; Vector: pRT3D-Pac; Site 1: EcoRI; Site 2: NotI"

BASE COUNT 65 a 51 c 72 g 78 t
 ORIGIN

Query Match 13.1%; Score 175; DB 9; Length 266;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 225; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 945 CAGTTACGAGAGATTATGCTGCGACCCCTACACACACACTTGTCTCAGCCCCAC 1004
 DB 246 CAGTTACGAGAGATTATGCTGCGACCCCTACACACACACTTGTCTCAGCCCCAC 187

OY 1005 CTACGCGCTGTGTCATGAATGCTTTTGCGCCCTTGACCGATCCCAAGATAGAGCCA 1064
 DB 186 CTACGCGCTGTGTCATGAATGCTTTTGCGCCCTTGACCGATCCCAAGATAGAGCCA 127

OY 1065 TGCTGATGATGAGGTCGCTCTTTCTTCTTCATTCAGAGCTAGTATATCAAGGGGAGTA 1124
 DB 126 TGCTGATGATGAGGTCGCTCTTTCTTCTTCATTCAGAGCTAGTATATCAAGGGGAGTA 67

OY 1125 CAACCGTTTGTCTCCATATTAATGATTAACCATTAACCAACAA 1170
 DB 66 CAACCGTTTGTCTCCATATTAATGATTAACCATTAACCAACAA 21

RESULT 7
 LOCUS A1850237
 DEFINITION UI-M-BG1-a1e-e-06-0-UI.s1 NIH BMAP MSC N Mus musculus cDNA clone UI-M-BG1-a1e-e-06-0-UI 3', mRNA sequence.
 ACCESSION A1850237
 VERSION A1850237.1 GI:5494143
 KEYWORDS EST.
 SOURCE house mouse.

Oligo-dt track not found, Not I site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M. Facina Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics This clone is also available through the I.M.G.E. Consortium at LNT (infoimage.lnt.gov). IMAGE ID=177166
Seq primer: M13 Forward
POLYA=No.

FEATURES

source

1. 460
Location/Qualifiers
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-E0-bv-c-04-0-UI"
/clone_lib="UI-R-E0"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: NotI; Site 2: EcoRI. This library consists of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dt track which allows identification of the library of origin of a clone within the mixture."

BASE COUNT 119 a 156 c 100 g 85 t

ORIGIN

Query Match 19.1%; Score 256; DB 9; Length 460;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 CCGTCTCCGGACCCGACACAGATGATGCGGCCGACGACGCGCCGACG 259

DB 205 CCGTCTCCGGACCCGACACAGATGATGCGGCCGACGACGCGCCGACG 264

QY 260 CACACCTTTGAAAACACAGAAAAGTCCAGCCAGCCGCTGATGTTCACA 319

DB 265 CACACCTTTGAAAACACAGAAAAGTCCAGCCAGCCGCTGATGTTCACA 324

QY 320 TCCCTTCGGGTTCCGGGATCCAGACCTCCGACAAATGTTGGCAATTTGTA 379

DB 325 TCCCTTCGGGTTCCGGGATCCAGACCTCCGACAAATGTTGGCAATTTGTA 384

QY 380 TAGATGTTGAATTTATTTATGAGCGGGCTCGAAGGATTTGTTCTTAATTT 439

DB 385 TAGATGTTGAATTTATTTATGAGCGGGCTCGAAGGATTTGTTCTTAATTT 444

QY 440 AAAATAGTGGGATGC 455

DB 445 AAAATAGTGGGATGC 460

RESULT 4
A1501505 324 bp mRNA linear EST 11-MAR-1999

LOCUS UI-R-C2p-rz-c-09-0-UI.s1 UI-R-C2p Rattus norvegicus cDNA clone
DEFINITION UI-R-C2p-rz-c-09-0-UI 3', mRNA sequence.

ACCESSION A1501505
VERSION A1501505.1 GI:4399356
KEYWORDS EST.

SOURCE Norway rat.
Rattus norvegicus

ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 324)

AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

COMMENT Contact: Soares, MB

Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu

Oligo-dt track not found, Not I site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)

FEATURES

source

1. 324
Location/Qualifiers

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C2p-rz-c-09-0-UI"
/clone_lib="UI-R-C2p"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI. The UI-R-C2p library is a subtracted library derived from the UI-R-C1 library, which is a subtracted library derived from the UI-R-C0 library. The UI-R-C0 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dt track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C2p) was constructed as follows: PCR amplified cDNA inserts from UI-R-C1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C2p library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"

BASE COUNT 88 a 116 c 64 g 56 t

ORIGIN

Query Match 15.0%; Score 201; DB 9; Length 324;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 CCGTCTCCGGACCCGACACAGATGATGCGGCCGACGACGCGCCGACG 259

DB 124 CCGTCTCCGGACCCGACACAGATGATGCGGCCGACGACGCGCCGACG 183

QY 260 CACACCTTTGAAAACACAGAAAAGTCCAGCCAGCCGCTGATGTTCACA 319

DB 184 CACACCTTTGAAAACACAGAAAAGTCCAGCCAGCCGCTGATGTTCACA 243

QY 320 TCCCTTCGGGTTCCGGGATCCAGACCTCCGACAAATGTTGGCAATTTGTA 379

DB 244 TCCCTTCGGGTTCCGGGATCCAGACCTCCGACAAATGTTGGCAATTTGTA 303

QY 380 TAGATGTTGAATTTATTTTA 400

DB 304 TAGATGTTGAATTTATTTTA 324

RESULT 5
A1843387 379 bp mRNA linear EST 14-JUL-1999

LOCUS A1843387/c
DEFINITION UI-M-AQ1-ad2-f-10-0-UI.s1 NIH_BMAP_MRI_N Mus musculus cDNA clone

ACCESSION UI-M-AQ1-ad2-f-10-0-UI 3', mRNA sequence.

QY 380 TAGATGTTGAATTTATTTTA 400

DB 304 TAGATGTTGAATTTATTTTA 324

QY 380 TAGATGTTGAATTTATTTTA 400

DB 304 TAGATGTTGAATTTATTTTA 324

QY 380 TAGATGTTGAATTTATTTTA 400

DB 304 TAGATGTTGAATTTATTTTA 324

QY 380 TAGATGTTGAATTTATTTTA 400

DB 304 TAGATGTTGAATTTATTTTA 324

```

/db_xref="ATCC (inhost):2022664"
/db_xref="taxon:10118"
/clone="RBRK70"
/clone_lib="Normalized rat brain, Bento Soares"
/notes="Organ: brain; Vector: pT73Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT      146 a 168 c 128 g 100 t
ORIGIN
Query Match      25.7%; Score 344; DB 9; Length 542;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 344; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 200 CCGTCTCCGGACCCGACACAGACAGATGATGCCGCCGACGCGCCAGCCCGCAGA 259
Db 199 CCGTCTCCGGACCCGACACAGACAGATGATGCCGCCGACGCGCCAGCCCGCAGA 258

Qy 260 CACAACCTTCTGAAAAACAGAAAAAAGTCCAGCCCAAGCGGCTGCATGTGTCAACA 319
Db 259 CACAACCTTCTGAAAAACAGAAAAAAGTCCAGCCCAAGCGGCTGCATGTGTCAACA 318

Qy 320 TCCCTTCGGTTCGGGATCCAGACTCCGCAAAATGTTGGCCAAATTTGGTAAATAT 379
Db 319 TCCCTTCGGTTCGGGATCCAGACTCCGCAAAATGTTGGCCAAATTTGGTAAATAT 378

Qy 380 TAGATGTTGAATATTTTAAATAGAGCGGGCTCGAAGGGATTTGGTTCTTAACCTTTG 439
Db 379 TAGATGTTGAATATTTTAAATAGAGCGGGCTCGAAGGGATTTGGTTCTTAACCTTTG 438

Qy 440 AAAATAGTCCGATCCGACAGGGCGAGGAGAAATTCGACGGTACCGTGTAGAGGCC 499
Db 439 AAAATAGTCCGATCCGACAGGGCGAGGAGAAATTCGACGGTACCGTGTAGAGGCC 498

Qy 500 GTAATAGAGGTTAATATGAGCAGACGCGGTATGACTAAT 543
Db 499 GTAATAGAGGTTAATATGAGCAGACGCGGTATGACTAAT 542

RESULT 2
BF420112      494 bp      mRNA      linear      EST 28-NOV-2000
LOCUS         UI-R-BJ2-bpJ-a-08-0-UI.s1 UI-R-BJ2 Rattus norvegicus cDNA clone
DEFINITION   UI-R-BJ2-bpJ-a-08-0-UI 3', mRNA sequence.
ACCESSION    BF420112
VERSION      BF420112.1 GI:11408101
KEYWORDS     EST.
SOURCE       Norway rat.
ORGANISM     Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE    1 (bases 1 to 494)
AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE       Normalization and subtraction: two approaches to facilitate gene
JOURNAL      Genome Res. 6 (9), 791-806 (1996)
MEDLINE      97044477
COMMENT      Contact: Soares, MB
              Program for Rat Gene Discovery and Mapping
              University of Iowa
              451 Eckstein Medical Research Building Iowa City, IA 52242, USA
              Tel: 319 335 8250
              Fax: 319 335 9565
              Email: msoares@iue.weeg.uiowa.edu
              Oligo-dT track not found, Not I site shown in beginning of sequence
              is likely internal to the message. cDNA Library Preparation: M.B.
              Soares Lab Clone distribution: clones will be available through
              Research Genetics (www.resgen.com)
              Seq primer: M13 Forward
              POLYA=No
              Location/Qualifiers
              1..494
              /organism="Rattus norvegicus"

```

```

/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ2-bpJ-a-08-0-UI"
/clone_lib="UI-R-BJ2"
/db_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ2
library is a subtracted library derived from the following
tissues: heart, atrium at 15 dpc, ventricle at 16.5 dpc,
atrium at 16.5 dpc, ventricle at 13 dpc, ventricle at 15
dpc. AV canal at 15 dpc. For a detailed description of
the library from which this clone was derived, please
visit our web site at latest.eng.uiowa.edu. The
subtraction has been previously described in (Bonaldo,
Lennon and Soares, Genome Research 6:791-806, 1996)
TAG SEQ=None found"
BASE COUNT      128 a 162 c 115 g 89 t
ORIGIN
Query Match      21.6%; Score 289; DB 12; Length 494;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 289; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 200 CCGTCTCCGGACCCGACACAGACAGATGATGCCGCCGACGCGCCAGCCCGCAGA 259
Db 206 CCGTCTCCGGACCCGACACAGACAGATGATGCCGCCGACGCGCCAGCCCGCAGA 265

Qy 260 CACAACCTTCTGAAAAACAGAAAAAAGTCCAGCCCAAGCGGCTGCATGTGTCAACA 319
Db 266 CACAACCTTCTGAAAAACAGAAAAAAGTCCAGCCCAAGCGGCTGCATGTGTCAACA 325

Qy 320 TCCCTTCGGTTCGGGATCCAGACTCCGCAAAATGTTGGCCAAATTTGGTAAATAT 379
Db 326 TCCCTTCGGTTCGGGATCCAGACTCCGCAAAATGTTGGCCAAATTTGGTAAATAT 385

Qy 380 TAGATGTTGAATATTTTAAATAGAGCGGGCTCGAAGGGATTTGGTTCTTAACCTTTG 439
Db 386 TAGATGTTGAATATTTTAAATAGAGCGGGCTCGAAGGGATTTGGTTCTTAACCTTTG 445

Qy 440 AAAATAGTCCGATCCGACAGGGCGAGGAGAAATTCGACGGTACCGT 488
Db 446 AAAATAGTCCGATCCGACAGGGCGAGGAGAAATTCGACGGTACCGT 494

RESULT 3
AA859573      460 bp      mRNA      linear      EST 03-JUL-1999
LOCUS         UI-R-E0-DV-C-04-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone
DEFINITION   UI-R-E0-DV-C-04-0-UI 3' similar to emb|AL009266|HSP38A20 Homo
              sapiens cDNA similar to RNA binding protein C. elegans, complete,
              mRNA sequence.
ACCESSION    AA859573
VERSION      AA859573.1 GI:4230119
KEYWORDS     EST.
SOURCE       Norway rat.
ORGANISM     Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE    1 (bases 1 to 460)
AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE       Normalization and subtraction: two approaches to facilitate gene
JOURNAL      Genome Res. 6 (9), 791-806 (1996)
MEDLINE      97044477
COMMENT      On Mar 10, 1998 this sequence version replaced gi:2949093.
              Contact: Soares, MB
              Program for Rat Gene Discovery and Mapping
              University of Iowa
              451 Eckstein Medical Research Building Iowa City, IA 52242, USA
              Tel: 319 335 8250
              Fax: 319 335 9565
              Email: msoares@iue.weeg.uiowa.edu

```

GenCore version 5.1.4_j05_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 15:59:00 ; Search time 1885.21 Seconds
(without alignments)
11511.672 Million cell updates/sec

Title: US-09-809-545A-1

Perfect score: 1340
Sequence: 1 ggcggccggcccgacacaa.....ggccgaagggtcgcagaa 1340

Scoring table: OLIGO_NUC
Gapco 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST: *
1: em_gesba: *
2: em_gesbun: *
3: em_geslin: *
4: em_geslun: *
5: em_gesrov: *
6: em_gesrpl: *
7: em_gestro: *
8: em_hcc: *
9: gb_esc1: *
10: gb_esc2: *
11: gb_hcc: *
12: gb_esc3: *
13: gb_esc4: *
14: gb_esc5: *
15: em_escfun: *
16: em_escrom: *
17: gb_gesb: *
18: em_gesb_hum: *
19: em_gesb_inv: *
20: em_gesb_pln: *
21: em_gesb_vtc: *
22: em_gesb_fun: *
23: em_gesb_man: *
24: em_gesb_mus: *
25: em_gesb_other: *
26: em_gesb_pro: *
27: em_gesb_tod: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	344	25.7	542	9	A1101278 EST210567
2	289	21.6	494	12	BF420112 UT-R-BJ2-B
3	256	19.1	460	9	AA859573 UT-R-BO-b
4	201	15.0	324	9	A1501505 UT-R-C2P-
5	183	13.7	379	9	A1843387 UT-M-AQ1-
6	175	13.1	266	9	A1175239 EST218774

Result No.	Score	Query Match	Length	DB ID	Description
7	175	13.1	430	9	A1850237 UT-M-BG1-
8	168	12.5	538	9	AA818804 UT-R-A0-a
9	160	11.9	533	12	BF549922 UT-R-E0-b
10	157	11.7	361	9	A1835011 UT-M-AM1-
11	146	10.9	425	10	BB750925 BB750925
12	140	10.4	395	9	A1717457 UT-R-Y0-a
13	97	7.2	600	13	B188857 4023-45 M
14	80	6.0	408	12	BG795922 UTSM-SM38
15	70	5.2	228	14	Z28900 HSBH7F031 S
16	70	5.2	327	12	BG186689 RST5664 A
17	70	5.2	363	14	D60520 HUM114E06A
18	70	5.2	367	9	A1095813 CB20911.x
19	70	5.2	383	9	A1189407 qd01805.x
20	70	5.2	391	9	A1186273 qd20808.x
21	70	5.2	392	9	AA975235 cq36c08.s
22	70	5.2	420	9	A1655094 wD67a10.x
23	70	5.2	427	12	BF223478 7G33601.x
24	70	5.2	439	9	A1968176 wu14a06.x
25	70	5.2	439	10	AM889795 hg22d04.x
26	70	5.2	439	10	BE501380 7a41h09.x
27	70	5.2	443	9	AA773715 aE81C05.T
28	70	5.2	449	9	A1374891 ta6012.x
29	70	5.2	456	9	A1917298 ts96809.x
30	70	5.2	459	9	A1244212 q386b11.x
31	70	5.2	477	10	AM197589 xm44g02.x
32	70	5.2	480	9	A1401040 t127a12.x
33	70	5.2	495	9	A1656926 tE48e08.x
34	70	5.2	607	9	AU145000 AU145000
35	70	5.2	773	10	AV729198 AV729198
36	70	5.2	774	10	AV729057 AV729057
37	70	5.2	4039	11	BC026312 Homo. gap1
38	66	4.9	283	12	BG185609 RST4560 A
39	62	4.6	674	12	BG793800 UTSW-SM16
40	55	4.1	228	10	AV339026 AV339026
41	55	4.1	228	10	AV349493 AV349493
42	55	4.1	226	9	AV253203 AV253203
43	55	4.1	274	10	BB197021 BB197021
44	55	4.1	277	10	BB431921 BB431921
45	55	4.1	295	10	BB188777 BB188777

ALIGNMENTS

RESULT 1
A1101278 542 bp mRNA linear EST 31-JAN-1999
DEFINITION
RBRER70 3' end, mRNA sequence.
ACCESSION
A1101278
VERSION
A1101278.1 GI:3704621
KEYWORDS
EST.
SOURCE
Rattus sp.
ORGANISM
Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE
1 (bases 1 to 542)
Lee,N.H., Glodak,A., Chandra,I., Mason,T.M., Quackenbush,J., Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat Gene Index
Unpublished (1998)
CONTACT: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@igr.org
Seq primer: M13-21.
LOCATION/Qualifiers
1..542
/organism="Rattus sp."

CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AA564197-AA594564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC
SQ Sequence 2118 BP; 557 A; 516 C; 542 G; 503 T; 0 other;
Query Match 2.9%; Score 39; DB 23; Length 2118;
Best Local Similarity 100.0%; Pred. No. 66-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 946 AGTTACGACGAGTTATGCTGCCGACCCCTACCCAC 984
Db 1300 AGTTACGACGAGTTATGCTGCCGACCCCTACCCAC 1338
RESULT 14
AAH03678
ID AAH03678 standard; cDNA; 893 BP.
AC AAH03678;
XX
XX 26-JUN-2001 (first entry)
DT
XX
XX Human cDNA clone (5'-primer) SEQ ID NO:513.
DE
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
OS
XX Homo sapiens.
XX
XX EP1074617-A2.
PN
XX
XX 07-FEB-2001.
PD
XX
XX 28-JUL-2000; 2000EP-0116126.
PF
XX
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
PA
XX
XX Ota T, Isegai T, Nishikawa T, Hayaishi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Makamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
DR
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX
XX Claim 1; SEQ ID 513; 2537bp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-CT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH1633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH16332
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
CC
SQ Sequence 893 BP; 225 A; 228 C; 236 G; 199 T; 5 other;
Query Match 2.6%; Score 35; DB 22; Length 893;
Best Local Similarity 100.0%; Pred. No. 0.00026;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 415 AAGGATTGCTTCGTAACCTTCGAAATAGTGC 449
Db 686 AAGGATTGCTTCGTAACCTTCGAAATAGTGC 720
RESULT 15
AAS31173/C
ID AAS31173 standard; cDNA; 1295 BP.
XX
XX AAS31173;
AC
XX
XX 04-DEC-2001 (first entry)
DT
XX
XX
XX Human diagnostic and therapeutic polynucleotide (DITHP) #188.
DE
XX
XX Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
KW cell proliferative disorder; Crohn's disease; lymphoma; leukemia;
KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;
KW respiratory disorder; ss.
XX
XX
XX Homo sapiens.
OS
XX
XX WO200162927-A2.
PN
XX
XX 30-AUG-2001.
PD
XX
XX 21-FEB-2001; 2001WO-US06059.
PF
XX
XX 24-FEB-2000; 2000US-0184693.
PR 24-FEB-2000; 2000US-0184697.
PR 24-FEB-2000; 2000US-0184698.
PR 24-FEB-2000; 2000US-0184768.
PR 24-FEB-2000; 2000US-0184769.
PR 24-FEB-2000; 2000US-0184770.
PR 24-FEB-2000; 2000US-0184771.
PR 24-FEB-2000; 2000US-0184772.
PR 24-FEB-2000; 2000US-0184773.
PR 24-FEB-2000; 2000US-0184774.
PR 24-FEB-2000; 2000US-0184776.
PR 24-FEB-2000; 2000US-0184777.
PR 24-FEB-2000; 2000US-0184797.
PR 24-FEB-2000; 2000US-0184813.
PR 24-FEB-2000; 2000US-0184837.
PR 24-FEB-2000; 2000US-0184841.
PR 24-FEB-2000; 2000US-0185213.
PR 24-FEB-2000; 2000US-0185216.
PR 12-MAY-2000; 2000US-0203785.
PR 15-MAY-2000; 2000US-0204226.
PR 16-MAY-2000; 2000US-0204525.
PR 16-MAY-2000; 2000US-0204821.
PR 16-MAY-2000; 2000US-0204908.
PR 16-MAY-2000; 2000US-0205232.
PR 17-MAY-2000; 2000US-0204815.
PR 17-MAY-2000; 2000US-0204863.
PR 17-MAY-2000; 2000US-0205221.
PR 17-MAY-2000; 2000US-0205285.
PR 17-MAY-2000; 2000US-0205286.
PR 17-MAY-2000; 2000US-0205287.
PR 17-MAY-2000; 2000US-0205323.

CC conditions (e.g. asthma). They are also useful for treating
CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
CC useful for tissue regeneration, for wound healing and in the treatment
CC of burns, incisions and ulcers. The proteins are also useful for
CC regulating haematopoiesis, for treating myeloid or lymphoid cell
CC deficiencies. The present sequence is one of the 625 cDNA sequences
CC encoding a secreted protein.

XX
SQ Sequence 3189 BP; 1005 A; 554 C; 631 G; 999 T; 0 other;

Query Match 3.5%; Score 47; DB 24; Length 3189;
Best Local Similarity 100.0%; Pred. No. 4.1e-09;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1068 TGATGATGTGGGTCTCGTCTTTCTTCTTCTTCATTCAGGCTAGTATATACC 1114
DB 515 TGATGATGTGGGTCTCGTCTTTCTTCTTCATTCAGGCTAGTATATACC 561

RESULT 12
AAK53229/c
ID AAK53229 standard; cDNA; 1164 BP.

XX AAK53229;

DT 06-NOV-2001 (first entry)

XX Human polynucleotide SEQ ID NO 2758.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukemia;
KW nervous system disorder; arthritis; inflammation; ss.

XX Homo sapiens.

OS WO200157190-A2.

XX PD 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-US04098.

XX PR 03-FEB-2000; 2000US-0496914.

XX PR 27-APR-2000; 2000US-0560875.

XX PR 20-JUN-2000; 2000US-0598075.

XX PR 19-JUL-2000; 2000US-0620325.

XX PR 01-SEP-2000; 2000US-0654936.

XX PR 15-SEP-2000; 2000US-0663561.

XX PR 20-OCT-2000; 2000US-0693325.

XX PR 30-NOV-2000; 2000US-0728422.

XX (HYSEQ-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao OA, Wang J, Zhang J, Ren F, Chen R, Wang ZM;
PI Xue AD, Yang Y, Wehrman T, Goodrich R;

XX WPI; 2001-476283/51.

XX P-PSDB; AAM80096.

XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -

XX Claim 1; Page 4968; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,

CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.

XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.

XX SQ Sequence 1164 BP; 273 A; 280 C; 325 G; 286 T; 0 other;

Query Match 3.2%; Score 43; DB 22; Length 1164;
Best Local Similarity 100.0%; Pred. No. 1.8e-07;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 634 CTGTTGTCCAGGCCAACAGAGGATCTTCCATGTACAGTG 676
DB 173 CTGTTGTCCAGGCCAACAGAGGATCTTCCATGTACAGTG 131

RESULT 13
AAS93634

XX ID AAS93634 standard; cDNA; 2118 BP.

XX AAS93634;

DT 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #29438.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; diagnostic disorder; ss.

XX Homo sapiens.

OS WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX (HYSEQ-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG29447.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX Claim 1; SEQ ID NO 29438; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.

XX The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity

```

XX  Sequence 2372 BP; 548 A; 726 C; 646 G; 452 T; 0 other;
SQ
Query Match          5.2%; Score 70; DB 21; Length 2372;
Best Local Similarity 100.0%; Pred. No. 4e-18;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

OY  1045 GATGCCAGACATGAGAGCGATCTATATATGAGGCTCTCTTCATTCATTCAGAGCT 1104
DB  2133 GATGCCAGACATGAGAGCGATCTATATATGAGGCTCTCTTCATTCATTCAGAGCT 2192
OY  1105 AGTATATACC 1114
DB  2193 AGTATATACC 2202

RESULT 10
ABN34672
ID  ABN34672 standard; DNA; 60 BP.
AC  ABN34672;
XX
XX  15-JUL-2002 (first entry)
DE  Human spliced transcript detection oligonucleotide SEQ ID NO:7420.
KW  Human; mouse; rat; splice transcript; detection; RNA transcript;
XX  splice variant; transcriptome; oligonucleotide library; ss.
OS  Homo sapiens.
XX
XX  W0200210449-A2.
PN  07-FEB-2002.
XX
XX  20-JUL-2001; 2001WO-1B01903.
PF  28-JUL-2000; 2000US-221607P.
PR  02-MAY-2001; 2001US-287724P.
XX
XX  (COMP-) COMPUGEN INC.
PA
PI  Shoshan A, Waseerman A, Mintz E, Mintz L, Faigler S;
XX
XX  WPI; 2002-257383/30.
DR
XX
XX  New oligonucleotide libraries comprising oligonucleotides which
PT  selectively hybridize to mRNAs transcribed from a transcription unit of
PT  a genome, useful for detecting tissue-, pathology-, and
PT  developmental-specific genes -
XX
XX  Example 1; SEQ ID 7420; 47p; English.
PS
XX
CC  The present invention describes oligonucleotide libraries for detecting
CC  messenger RNAs that populate a (sub-)transcriptome, where the
CC  (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC  transcription units that populate a genome. The library comprises
CC  several oligonucleotides, each capable of hybridizing selectively to a
CC  set of messenger RNAs transcribed from a given transcription unit of
CC  the genome, which encodes one or more messenger RNA splice variants.
CC  The oligonucleotide libraries are useful for detecting mRNAs from a
CC  biological sample, in expression profiling studies, in qualitatively or
CC  quantitatively characterizing the corresponding transcriptome, and in
CC  detecting RNA transcripts and splice variants of human or animal
CC  transcriptomes. The libraries may also be used as specialised mini
CC  libraries to detect transcripts of a sub-transcriptome under a
CC  particular biological or pathological state, and so allowing the
CC  detection of tissue- and pathology-specific genes such as those genes
CC  only expressed in specific tissue under a specific pathological
CC  condition, to detect developmental specific genes; and to detect RNA
CC  transcripts and splice variants of a transcriptome of a patient suffering
CC  from a particular disorder. ABN27253 to ABN59589 represent
CC  oligonucleotide sequences from rats, humans and mice, which are used in

```

the exemplification of the present invention.

N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 60 BP; 11 A; 14 C; 16 G; 19 T; 0 other;

Query Match 3.6%; Score 48; DB 24; Length 60;
Best Local Similarity 100.0%; Pred. No. 3e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1045 GATGCCAAGACTAGAGCCATGTCGATGATGCGGTCCTGTTCTTCT 1092
|||||
13 GATGCCAAGACTAGAGCCATGTCGATGATGCGGTCCTGTTCTTCT 60

RESULT 11
ABK34602
ID ABK34602 standard; cDNA; 3189 BP.
XX
AC ABK34602;
XX
DT 08-MAY-2002 (first entry)
XX
DE Human cDNA for novel secreted protein, SEQ ID 371.
XX
KW Human; ss; gene; secreted protein; immune deficiency; viral infection;
KW bacterial infection; fungal infection; autoimmune disorder; burn;
KW rheumatoid arthritis; multiple sclerosis; autoimmune thyroiditis;
KW diabetes; allergy; asthma; neurodegenerative disease; Crohn's disease;
KW Alzheimer's disease; Parkinson's disease; liver fibrosis; tumour;
KW coagulation disorder; haemophilia; inflammatory disorder; ulcer;
KW tissue regeneration; wound healing; haematopoiesis; myeloid deficiency;
KW lymphoid cell deficiency.
XX
OS Homo sapiens.
XX
XX Homo sapiens.
XX WO200177290-A2.
XX
XX 18-OCT-2001.
XX
XX 29-MAR-2001; 2001MO-US10295.
XX
XX 06-APR-2000; 2000US-194941P.
XX
XX (GEMV) GENETICS INST INC.
XX
XX Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
XX Gutukota K, Graham JR;
XX
XX WPI; 2002-179323/23.
XX
XX Six hundred and twenty five polynucleotides derived from a variety of
XX human tissue sources which encode secreted proteins, useful for
XX treating immune deficiencies and disorders such as autoimmune disorders
XX
XX
XX Claim 1; Page 188-189; 339pp; English.
XX
XX The invention relates to 625 polynucleotides which have been derived from
XX a variety of human tissue sources and which encode novel secreted
XX proteins, their complements and sequences that hybridise to them.
XX Also included are a vector comprising the polynucleotide, a host cell
XX transformed with the vector, the proteins encoded by the
XX polynucleotides, antihodds that bind to the proteins and identification
XX of modulators of the proteins or the expression of the polynucleotide.
XX The polynucleotides can be used as probes for the identification
XX and isolation of full length cDNA and genomic DNA. The polynucleotides
XX and proteins can also be used as nutritional supplements. The protein
XX is useful in the treatment of various immune deficiencies and disorders
XX such as viral infections, bacterial infections, fungal infections,
XX autoimmune disorders (e.g. rheumatoid arthritis, multiple sclerosis,
XX autoimmune thyroiditis and diabetes) and allergic reactions and

XX AAK52245;
 AC ID
 DT 06-NOV-2001 (first entry)
 DE Human polynucleotide SEQ ID NO 790.
 XX
 DE Human polynucleotide SEQ ID NO 790.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation; ss.
 XX
 OS Homo sapiens.
 XX
 PN MO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US04098.
 XX
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0588075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 XX
 PA (HYSEQ-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 PI Zhao QA, Wang D, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
 DR WPI; 2001-476283/51.
 DR P-PSDB; AAM79112.
 XX
 PT Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX
 PS Claim 1; Page 2643-2645; 6221pp; English.
 XX
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAW80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 XX
 SQ Sequence 1800 BP; 432 A; 552 C; 481 G; 335 T; 0 other;
 Query Match 5.2%; Score 70; DB 22; Length 1800;
 Best Local Similarity 100.0%; Pred. No. 4.2e-18;
 Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1045 GATGCCAAGACTAGAGACCATGCTGATGATGCGGTCTGTTCTTTCATTGACGGCT 1104
 DB 1600 GATGCCAAGACTAGAGACCATGCTGATGATGCGGTCTGTTCTTTCATTGACGGCT 1659
 QY 1105 AGTATATACC 1114
 DB 1660 AGTATATACC 1669

RESULT 9
 AAA07075
 ID AAA07075 standard; cDNA; 2372 BP.
 XX
 AC AAA07075;
 XX
 DT 03-JUL-2000 (first entry)
 XX
 DE cDNA encoding human ataxin-2 binding protein (A2BP).
 XX
 KW Ataxin-2 binding protein; A2BP; human; RNA-binding; cell polarisation;
 KW neuronal plasticity; cellular degeneration signal transduction pathway;
 KW selective RNA transport; spinocerebellar ataxia type-2;
 KW hyperproliferative disorder; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 987..1979
 FT CDS /tag= a
 FT /product= "Human A2BP"
 XX
 PN MO200012710-A1.
 XX
 PD 09-MAR-2000.
 XX
 PF 01-SEP-1999; 99WO-US20156.
 XX
 PR 01-SEP-1998; 98US-0145391.
 XX
 PA (CEDA-) CEDARS SINAI MEDICAL CENT.
 XX
 PI Pulst SM, Shibata H;
 PI WPI; 2000-237873/20.
 DR P-PSDB; AAY81462.
 XX
 PT Nucleic acids encoding an ataxin-2 binding protein useful for
 PT inhibiting the expression of active proteins from the SCA2 gene for the
 PT treatment of spinocerebellar ataxia type-2 -
 XX
 PS Claim 6; Page 74-77; 82pp; English.
 XX
 CC This sequence represents cDNA encoding human ataxin-2 binding protein
 CC (A2BP). Nucleotide sequences encoding human A2BP were originally
 CC isolated in an adult brain cDNA library using the yeast two hybrid
 CC method. The ligand of A2BP, ataxin-2, is a 40 kD protein of unknown
 CC function that is encoded by the SCA2 gene located on chromosome 12. SCA2
 CC has been linked to the autosomal dominant neurodegenerative disorder
 CC spinocerebellar ataxia type-2. Individuals afflicted with the disease
 CC exhibit CAG triplet expansion in the SCA2 gene, resulting in ataxin-2
 CC containing a polyglutamine stretch of about 35-39 residues, whereas that
 CC of normal individuals contains approximately 22 contiguous glutamine
 CC residues. A2BP and ataxin-2 are components of a cellular degeneration
 CC signal transduction pathway. The pathogenic expanded form of ataxin-2 has
 CC a higher affinity for A2BP relative to normal ataxin-2; the presence of
 CC the expanded form is likely to promote degeneration. A2BP and ataxin have
 CC also been found to have a role in gene regulation. The binding of A2BP to
 CC ataxin-2 plays an important role in controlling gene expression via the
 CC targeting of transport of specific RNAs, selective RNA transport being
 CC mediated via the RNA binding domains of A2BP. A2BP is expressed very
 CC early in embryonic development. Both ataxin-2 and A2BP are able to bind
 CC RNA, in an essential components of the RNA localisation network that
 CC establishes cellular polarity in embryogenesis. In highly differentiated,
 CC polarised cells such as neurons, A2BP and ataxin-2 have a similar
 CC function and are required for neuronal plasticity. A2BP nucleic acids may
 CC be used for the recombinant production of A2BP proteins or fragments
 CC thereof according to standard methodologies. For example, an A2BP protein
 CC with an ataxin-2 or RNA binding capability but no signal transduction
 CC function can be used as a dominant negative inhibitor of the cellular
 CC degeneration signal transduction pathway. A2BP proteins with a signal
 CC transduction function can be used to treat hyperproliferative disorders
 CC (e.g., cancer) via stimulation of the cellular degeneration pathway.

(HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;
WPI: 2001-639362/73.
P-PSDB: ABG29446.

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -

Claim 1; SEQ ID No 29437; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIP0 at ftp.wipo.int/pub/published_pct_sequences.

Sequence 481 BP; 100 A; 145 C; 123 G; 113 T; 0 other;

Query Match 5.2%; Score 70; DB 23; Length 481;
Best Local Similarity 100.0%; Pred. No. 5.1e-18;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1045 GATGCCAGACTAGAGACCATGCTGATGATGAGGCTGCTTTCTTCATTTGAGGCT 1104
|||
306 GATGCCAGACTAGAGACCATGCTGATGATGAGGCTGCTTTCTTCATTTGAGGCT 365
|||
1105 AGTATTATCC 1114
|||||
366 AGTATTATACC 375

RESULT 5
ID ABK34530/C
ABK34530 standard; CDNA, 539 BP.

ABK34530;
08-MAY-2002 (first entry)

Human CDNA for novel secreted protein, SEQ ID 299.

Human; ss; gene; secreted protein; immune deficiency; viral infection; bacterial infection; fungal infection; autoimmune disorder; burn; rheumatoid arthritis; multiple sclerosis; autoimmune thyroiditis; diabetes; allergy; asthma; neurodegenerative disease; Crohn's disease; Alzheimer's disease; Parkinson's disease; liver fibrosis; tumour; coagulation disorder; haemophilia; inflammatory disorder; ulcer; tissue regeneration; wound healing; haematopoiesis; myeloid deficiency; lymphoid cell deficiency.

Homo sapiens.

WO200177290-A2.

PD	18-OCT-2001.
XX	
Pf	29-MAR-2001; 2001MO-US10295.
XX	
PR	06-APR-2000; 2000US-194941P.
XX	
PA	(GEMV) GENETICS INST INC.
XX	
P1	Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ,
PI	Gulukota K, Graham JR;
XX	
DR	WP1; 2002-179323/23.
XX	
PS	Claim 1; Page 153; 33ppp; English.
XX	
CC	The invention relates to 625 polynucleotides which have been derived from
CC	a variety of human tissue sources and which encode novel secreted
CC	proteins, their complements and sequences that hybridise to them.
CC	Also included are a vector comprising the polynucleotide, a host cell
CC	transformed with the vector, the proteins encoded by the
CC	polynucleotides, antibodies that bind to the proteins and identification
CC	of modulators of the proteins or the expression of the polynucleotide.
CC	The polynucleotides can be used as probes for the identification
CC	and isolation of full length cDNA and genomic DNA. The polynucleotides
CC	and proteins can also be used as nutritional supplements. The protein
CC	is useful in the treatment of various immune deficiencies and disorders
CC	such as viral infections, bacterial infections, fungal infections,
CC	autoimmune disorders (e.g. rheumatoid arthritis, multiple sclerosis,
CC	autoimmune thyroiditis and diabetes) and allergic reactions and
CC	conditions (e.g. asthma). They are also useful for treating
CC	neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
CC	disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
CC	inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
CC	useful for tissue regeneration, for wound healing and in the treatment
CC	of burns, incisions and ulcers. The proteins are also useful for
CC	regulating hematopoiesis, for treating myeloid or lymphoid cell
CC	deficiencies. The present sequence is one of the 625 cDNA sequences
CC	encoding a secreted protein.
CC	
SQ	Sequence 539 BP; 132 A; 132 C; 156 G; 119 T; 0 other;
XX	
XX	
Query Match	5.2%; Score 70; DB 24; Length 539;
Best Local Similarity	100.0%; Pred. No. 5e-18;
Matches 70; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Oy	1045 GATGCCAGACTAGAGCGCATGTGATGTGGTCTGTTCTTTCATTGCAGGCT 1104 DbB 131 GATGCCAGACTAGAGCGCATGTGATGTGGTCTGTTCTTTCATTGCAGGCT 132
Oy	1105 AGTATATACC 1114 Db 131 AGTATATACC 122
RESULT 6	
ID	AAH09205/c
AAH09205 standard; cDNA; 607 BP.	
AC	AAH09205;
DT	26-JUN-2001 (first entry)
DE	Human cDNA clone (3'-primer) SEQ ID NO:6040.
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
OS	Homo sapiens

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 12:52:03 ; Search time 313.615 Seconds
(without alignments)
9625.317 Million cell updates/sec

Title: US-09-809-545A-1
Perfect score: 1340
Sequence: 1 gcggccgcgcgcgcgcacacat.....ggccgaaggcgctcgtcta 1340

Scoring table: OLIGO_NUC
Gapco 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :
1: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1980.DAT:*
2: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1981.DAT:*
3: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1982.DAT:*
4: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1983.DAT:*
5: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1984.DAT:*
6: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1985.DAT:*
7: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1986.DAT:*
8: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1987.DAT:*
9: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1988.DAT:*
10: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1989.DAT:*
11: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1990.DAT:*
12: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1991.DAT:*
13: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1992.DAT:*
14: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1993.DAT:*
15: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1994.DAT:*
16: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1995.DAT:*
17: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1996.DAT:*
18: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1997.DAT:*
19: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1998.DAT:*
20: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1999.DAT:*
21: /SID52/gcgdata/geneeq/geneeqn-emb1/NA2000.DAT:*
22: /SID52/gcgdata/geneeq/geneeqn-emb1/NA2001A.DAT:*
23: /SID52/gcgdata/geneeq/geneeqn-emb1/NA2001B.DAT:*
24: /SID52/gcgdata/geneeq/geneeqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1340	100.0	1340	24	AAS94693	Rat secreted factor
2	91	6.8	327	22	AAK54001	Murine transcript
3	70	5.2	406	23	AAS93632	DNA encoding novel
4	70	5.2	481	23	AAS93633	DNA encoding novel
5	70	5.2	539	24	ABK34530	Human CDNA for nov
6	70	5.2	607	22	AAH09205	Human CDNA clone (
7	70	5.2	1513	22	AAH13824	Human CDNA sequenc
8	70	5.2	1800	22	AAK52245	Human polynucleoti
9	70	5.2	2372	21	AAA07075	CDNA encoding huma

10	48	3.6	60	24	ABN34672	Human spliced tran
11	47	3.5	3189	24	ABK34602	Human CDNA for nov
12	43	3.2	1164	22	AAK53329	Human polynucleoti
13	39	2.9	2118	23	AAS93634	DNA encoding novel
14	35	2.6	893	22	AAH03678	Human CDNA clone (
15	35	2.6	1295	22	AAS31173	Human diagnostic a
16	32	2.4	125	32	AAK20380	Human brain expres
17	32	2.4	125	22	AAK46486	Human bone marrow
18	32	2.4	125	22	AAI52354	Probe #21040 used
19	32	2.4	125	24	ABK20830	Human genome-deriv
20	32	2.4	394	20	AAV87007	EST clone AY147.
21	32	2.4	473	22	AAK07716	Human brain expres
22	32	2.4	473	22	AAK33549	Human bone marrow
23	32	2.4	473	22	AAI39304	Probe #7990 used t
24	32	2.4	473	22	ABK08416	Human genome-deriv
25	32	2.4	593	23	ABV40823	Human prostate exp
26	32	2.4	511	23	ABV38877	Human prostate exp
27	30	2.2	979	24	ABK29216	Oligonucleotide fo
28	30	2.2	979	24	ABK29217	Oligonucleotide fo
29	30	2.2	1168	24	ABK47032	Oligonucleotide fo
30	30	2.2	1168	24	ABK47033	Oligonucleotide fo
31	30	2.2	1574	21	AAK18023	Lung cancer associ
32	30	2.2	2308	24	ABK39167	Oligonucleotide fo
33	30	2.2	2308	24	ABK39167	Oligonucleotide fo
34	30	2.2	2365	24	ABK17068	Oligonucleotide fo
35	30	2.2	2365	24	ABK17069	Oligonucleotide fo
36	30	2.2	2395	24	ABK13720	Oligonucleotide fo
37	30	2.2	2395	24	ABK13721	Oligonucleotide fo
38	30	2.2	5695	22	AAS46538	Tumour suppressor
39	30	2.2	5695	22	AAS46538	Tumour suppressor
40	30	2.2	5898	24	ABK33940	Chemically pretrea
41	30	2.2	5898	24	ABK28329	Human immune syste
42	30	2.2	6126	22	AAS46574	DNA transcripion
43	30	2.2	6126	24	ABK33831	Tumour suppressor
44	30	2.2	6154	24	ABK33305	Human immune syste
45	30	2.2	6510	22	AAS46647	Tumour suppressor

ALIGNMENTS

RESULT 1	AAS94693	Rat secreted factor DNA clone P0184_D11 #1.
XX	XX	AAK54693 :
XX	XX	12-MAR-2002 (first entry)
DE	XX	Rat secreted factor DNA clone P0184_D11 #1.
XX	XX	Rat secreted factor polypeptide; cardiac disease; kidney;
KW	XX	inflammatory disease; congestive heart failure; myocarditis; asthma; ss;
KW	XX	dilated congestive cardiomyopathy; angina pectoris; cardiac arrhythmia;
KW	XX	myocardial infarction; pulmonary hypertension; arteriosclerosis; stroke;
KW	XX	atherosclerosis; cardiac tumour; glomerulonephritis; nephrotic syndrome;
KW	XX	renal infarction; hereditary nephritis; polycystic kidney disease;
KW	XX	chronic renal failure; renal vein thrombosis; medullary sponge kidney;
KW	XX	rheumatoid arthritis; osteoarthritis; psoriasis; restenosis; PCR primer;
KW	XX	graft versus host reaction; Crohn's disease; ulcerative colitis; probe;
KW	XX	Alzheimer's disease; gene therapy.
XX	XX	Rattus norvegicus.
OS	XX	
PN	XX	W0200174901-A2.
PD	XX	11-OCT-2001.
XX	XX	
PF	XX	23-MAR-2001; 2001WO-US09555.
XX	XX	
PR	XX	31-MAR-2000; 2000US-193548P.
XX	XX	14-MAR-2001; 2001US-0809545.

THIS PAGE BLANK (U.S.)

	Matches	67;	Conservative	.0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	881	AGGGTGGTTACGCTGCATACCGCTACGCCCGCAGCCCGCCACTGCTGCTGCCTACA								940
Db	34697	AGGGTGGTTACGCTGCATACCGCTACGCCCGCAGCCCGCCACTGCTGCTGCCTACA								34756
Qy	941	GTGACAG								947
Db	34757	GTGACAG								34763

Search completed: March 15, 2003, 22:48:10
Job time : 3992.54 secs

RESULT 15	AC120661	115860 bp	DNA	linear	HTG 23-JUL-2002
LOCUS	Rattus norvegicus clone CH230-34C2, *** SEQUENCING IN PROGRESS ***				
DEFINITION	21 unordered pieces.				
ACCESSION	AC120661.2	GI:21902877			
VERSION	HTG; HTGS_PHASE1.				
KEYWORDS	Rattus norvegicus				
SOURCE	Rattus norvegicus				
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
REFERENCE	1 (bases 1 to 115860)				
AUTHORS	Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Amarantunga, H.C., Are, J.R., Ayele, M., Banks, D., Barbara, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, T., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haviak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lied, C., Liu, J., Liu, W., Loulseghe, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Roife, M., Ruiz, S., Savary, G., Scher, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Umani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G., and Gibbs, R.				
TITLE	Direct Submission				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 115860)				
AUTHORS	Worley, K.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
REFERENCE	3 (bases 1 to 115860)				
AUTHORS	Worley, K.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (23-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
COMMENT	On Jul 18, 2002 this sequence version replaced gi:20514522. ----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/				
CONTACT	hgsc-help@bcm.tmc.edu				
PROJECT INFORMATION	Center project name: GxLC Center clone name: CH230-34C2 Summary Statistics				
SEQUENCING VECTOR	Plasmid				
CHEMISTRY	Dye-terminator Big Dye; 100% of reads				
ASSEMBLY PROGRAM	Phrap; version 0.990329				
CONSENSUS QUALITY	99693 bases at least Q40				
CONSENSUS QUALITY	101936 bases at least Q30				
CONSENSUS QUALITY	102743 bases at least Q20				
NOTE	Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html). NOTE: This is a 'working draft' sequence. It currently consists of 21 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.				
FEATURES	Location/Qualifiers				
SOURCE	1. 115860 /organism="Rattus norvegicus" /db_xref="taxon:10116" /clone="CH230-34C2"				
BASE COUNT	33039 a 22417 c 22786 g 30574 t 7044 others				
ORIGIN					
QUERY MATCH	5.0%; Score 67; DB 2; Length 115860;				
BEST LOCAL SIMILARITY	100.0%; Pred. No. 1e-28;				

```
repeat_region complement(14642. .14755)
repeat_region /rpt_family="L1"
repeat_region 17429. .18370
repeat_region /rpt_family="MSTAR"
repeat_region complement(19470. .19819)
repeat_region /rpt_family="MLT1"
repeat_region complement(20352. .20627)
repeat_region /rpt_family="Alu"
repeat_region 23808. .24058
repeat_region /rpt_family="Alu"
repeat_region 27654. .28217
repeat_region /rpt_family="L1"
repeat_region complement(32764. .32956)
repeat_region /rpt_family="L1"
repeat_region complement(33392. .33503)
repeat_region /rpt_family="Alu"
repeat_region 34586. .34870
repeat_region /rpt_family="Alu"
repeat_region 34936. .35489
repeat_region /rpt_family="MLT2B2"
repeat_region complement(36940. .37172)
repeat_region /rpt_family="Alu"
repeat_region 37173. .37437
repeat_region /rpt_family="THE1"
repeat_region 41428. .41745
repeat_region /rpt_family="THE1"
repeat_region complement(42057. .43927)
repeat_region /rpt_family="L1"
repeat_region 43924. .44555
repeat_region /rpt_family="L1"
repeat_region 45128. .45367
repeat_region /rpt_family="Alu"
repeat_region 45467. .45704
repeat_region /rpt_family="Alu"
repeat_region 46193. .46578
repeat_region /rpt_family="MER25"
repeat_region complement(46779. .47141)
repeat_region /rpt_family="THE1"
repeat_region complement(48265. .48550)
repeat_region /rpt_family="Alu"
repeat_region 49630. .50221
repeat_region /rpt_family="MER41"
repeat_region 50781. .51096
repeat_region /rpt_family="Alu"
repeat_region complement(51219. .51439)
repeat_region /rpt_family="Alu"
repeat_region 51462. .51708
repeat_region /rpt_family="MER33"
repeat_region complement(52042. .52321)
repeat_region /rpt_family="Alu"
repeat_region 54001. .54092
repeat_region /rpt_family="Alu"
repeat_region complement(55760. .56014)
repeat_region /rpt_family="Alu"
repeat_region 56616. .56899
repeat_region /rpt_family="Alu"
repeat_region complement(66463. .66747)
repeat_region /rpt_family="Alu"
repeat_region complement(66908. .67180)
repeat_region /rpt_family="Alu"
repeat_region complement(70389. .70704)
repeat_region /rpt_family="MER1"
repeat_region complement(72166. .72421)
repeat_region /rpt_family="MER33"
repeat_region 72501. .72769
repeat_region /rpt_family="Alu"
repeat_region 72843. .72956
repeat_region /rpt_family="MER5"
repeat_region complement(76303. .76564)
repeat_region /rpt_family="Alu"
repeat_region 97534. .97700
repeat_region /rpt_family="MER3"
repeat_region complement(98469. .98526)
```

```
repeat_region /rpt_family="Alu"
repeat_region complement(99089. .99381)
repeat_region /rpt_family="Alu"
repeat_region 101648. .101930
repeat_region /rpt_family="Alu"
repeat_region 102790. .103056
repeat_region /rpt_family="Alu"
repeat_region 104010. .104277
repeat_region /rpt_family="Alu"
repeat_region complement(111121. .111263)
repeat_region /rpt_family="MERS"
repeat_region complement(112263. .112476)
repeat_region /rpt_family="MER30"
repeat_region 112956. .113234
repeat_region /rpt_family="Alu"
repeat_region 114385. .114657
repeat_region /rpt_family="Alu"
repeat_region complement(117031. .117194)
repeat_region /rpt_family="MER20"
repeat_region complement(118464. .118770)
repeat_region /rpt_family="Alu"
repeat_region 119271. .119549
repeat_region /rpt_family="Alu"
repeat_region complement(122831. .122940)
repeat_region /rpt_family="MER41"
repeat_region 123471. .123742
repeat_region /rpt_family="Alu"
repeat_region complement(129860. .129972)
repeat_region /rpt_family="Alu"
repeat_region 135235. .135310
repeat_region /rpt_family="MLT1"
repeat_region 135874. .136157
repeat_region /rpt_family="Alu"
repeat_region 139775. .140048
repeat_region /rpt_family="Alu"
repeat_region 141734. .141814
repeat_region /rpt_family="MIR"
repeat_region 143071. .143352
repeat_region /rpt_family="Alu"
repeat_region 148286. .148773
repeat_region /rpt_family="MER1"
repeat_region 150431. .150707
repeat_region /rpt_family="Alu"
repeat_region 155151. .155387
repeat_region /rpt_family="Alu"
repeat_region complement(156101. .156213)
repeat_region /rpt_family="MER20"
repeat_region complement(156636. .156920)
repeat_region /rpt_family="Alu"
repeat_region complement(159633. .159931)
repeat_region /rpt_family="Alu"
repeat_region complement(160378. .160658)
repeat_region /rpt_family="Alu"
repeat_region 161617. .162140
repeat_region /rpt_family="Alu"
repeat_region complement(163200. .163315)
repeat_region /rpt_family="MIR"

BASE COUNT 48114 a 34256 c 33738 g 48430 t
ORIGIN
```

Query Match 5.2%; Score 70; DB 9; Length 164538;
Best Local Similarity 100.0%; Pred.No. 1.5e-30;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1045 GATGCCAAGACTAGGAGCCATCTCATGTGCGTCTGCTCTTCTTCATTGCAGGCT 1104
|||||
DB 61328 GATGCCAAGACTAGGAGCCATCTCATGTGCGTCTGCTCTTCTTCATTGCAGGCT 61269
|||||

QY 1105 AGTATATACC 1114
|||||
DB 61268 AGTATATACC 61259
|||||

```

Db 180 GATGCCAAGACTAGGAGCCATGCTGATGGTCTGCTTTCTTCTTCATTGCAGGCT 239
|||||
Qy 1105 AGTATATACC 1114
|||||
Db 240 AGTATATACC 249
|||||

RESULT 13
AF109106 3348 bp mRNA linear PRI 01-MAR-2002
LOCUS Homo sapiens hexaribonucleotide binding protein 1 isoform beta
DEFINITION (HNRBP1) mRNA, complete cds.
ACCESSION AF109106
VERSION AF109106.1 GI:19032367
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3348)
AUTHORS Chen, W., Chu, Z.-L., Blough, R. I., Liu, L., Hoppes, B. and
Winkelmann, J. C.
TITLE Molecular Cloning and Chromosomal Localization of a Novel Human
Brain, Heart and Skeletal Muscle Specific RNA Binding Protein Gene
Homologous to fox-1 in Caenorhabditis elegans
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3348)
AUTHORS Chen, W., Chu, Z.-L., Blough, R. I., Liu, L., Hoppes, B. and
Winkelmann, J. C.
TITLE Direct Submission
JOURNAL Submitted (24-NOV-1998) Internal Medicine/Hematology-Oncology,
University of Cincinnati College of Medicine, 231 Bethesda Ave. ML
0508, Cincinnati, OH 45267-0508, USA
FEATURES
source
1..3348
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p13.3"
gene 1..3348
/genes="HNRBP1"
CDS 254..1441
/genes="HNRBP1"
/notes="similar
/codon_start=1
/product="hexaribonucleotide binding protein 1 isoform
beta"
/protein_id="AAL83406.1"
/db_xref="GI:19032368"
/translations="MLASQGVLLHPYGVPMIVPAAPYLPGLIQNGEAAAPDTMAQP
YASQAPPPONGIPAEYTAHPHPAPEYTGTTVPEHTLNLYPPAOTHSQSPADTSA
OTVSGTATDDAAPTQGPOTOPSENTENKSKPKRLHVSNIPEFRDPLROMFGQF
GKILDVILFNERGSKGFGFTVENSADAREKHLGTVVEGRKLEVNNAATAVMTN
KKTVNTYNGKLVNPGVAVSYPEFYAGTVLLCOANQEGSSMYSAFSLVYTSAMFGF
PYPAATAAAYRAGHRLGRGRTVNTFRAAAPPPPIPAYGGVYQEVYGNKLLQGGY
AAVRYAQTPTATAAAYSDRNQFVEVADEISCNSTSAVTDFMLPTPTTLLQPPPTA
LVP"
polyA_signal 3299..3304
/gene="HNRBP1"
BASE COUNT 950 a 775 c 711 g 912 t
ORIGIN
Query Match 5.2%; Score 70; DB 9; Length 3348;
Best Local Similarity 100.0%; Pred. No. 1.3e-30;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1045 GATGCCAAGACTAGGAGCCATGCTGATGGTCTGCTTTCTTCTTCATTGCAGGCT 1104
Db 1462 GATGCCAAGACTAGGAGCCATGCTGATGGTCTGCTTTCTTCTTCATTGCAGGCT 1521
Qy 1105 AGTATATACC 1114
|||||

```

```

Db 1522 AGTATATACC 1531
|||||
RESULT 14
AC005774/c 164538 bp DNA linear PRI 05-OCT-1998
LOCUS Homo sapiens chromosome 16, BAC clone 2603 (LANL), complete
DEFINITION sequence.
ACCESSION AC005774
VERSION AC005774.1 GI:3702264
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 164538)
AUTHORS Ricke, D. C.
TITLE Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 164538)
AUTHORS Ricke, D. C., Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E.,
Robinson, D., Jones, M., Buckingham, J., Chasteen, L., Thompson, S.,
Goodwin, L., Bryant, J., Tesmer, J., Meincke, L., Longmire, J.,
White, S., Ueng, S., Tatum, O., Campbell, C., Fawcett, J., Maltbie, M.
and Deaven, L.
TITLE Sequencing of Human Chromosome 16p13.3
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 164538)
AUTHORS Ricke, D. C., Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E.,
Robinson, D., Jones, M., Buckingham, J., Chasteen, L., Thompson, S.,
Goodwin, L., Bryant, J., Tesmer, J., Meincke, L., Longmire, J.,
White, S., Ueng, S., Tatum, O., Campbell, C., Fawcett, J., Maltbie, M.
and Deaven, L.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-1998) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
FEATURES
source
1..164538
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p13.3"
/clone="2603"
1378..1665
/rpt_family="Alu"
complement(2144..2320)
/rpt_family="MER21"
complement(2198..2416)
/rpt_family="MER39"
complement(2450..2725)
/rpt_family="Alu"
complement(3157..3393)
/rpt_family="Alu"
complement(3524..3792)
/rpt_family="Alu"
5731..6020
/rpt_family="Alu"
complement(7557..7727)
/rpt_family="MER4"
8282..8928
/rpt_family="LTR8"
10055..10310
/rpt_family="THE1"
10472..10618
/rpt_family="L1"
10839..10939
/rpt_family="Alu"
complement(11109..11357)
/rpt_family="Alu"
12050..12312
/rpt_family="Alu"

```

ACCESSION	gamma isoforms (HRNBPI) gene, alternatively spliced, exon 14 and partial cds and complete cds.
VERSION	AF109120
KEYWORDS	AF109120.1 GI:19032383
SEGMENT	14 of 14
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 2971)
TITLE	Chen,W. and Winkelmann,J.C.
JOURNAL	The Exon-Intron Organization of the Human HRNBPI Gene
REFERENCE	2 (bases 1 to 2971)
AUTHORS	Chen,W., Chu,Z.-L., Blough,R.I., Liu,L., Hoppes,B. and Winkelmann,J.C.
TITLE	Molecular Cloning and Chromosomal Localization of a Novel Human Brain, Heart and Skeletal Muscle Specific RNA Binding Protein Gene Homologous to fox-1 in Caenorhabditis elegans
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 2971)
AUTHORS	Chen,W. and Winkelmann,J.C.
TITLE	Direct Submission
JOURNAL	Submitted (25-NOV-1998) Internal Medicine/Hematology-Oncology, University of Cincinnati College of Medicine, 231 Bethesda Ave. ML 0508, Cincinnati, OH 45267-0508, USA
FEATURES	Location/Qualifiers
source	1..2971
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="16"
	/map="16p13.3"
gene	order(AF109107.1:372..948,AF109108.1:1..470,AF109109.1:1..529,AF109110.1:1..219,AF109111.1:1..219,AF109112.1:1..253,AF109113.1:1..347,AF109114.1:1..374,AF109115.1:1..251,AF109116.1:1..249,AF109117.1:1..151,AF109118.1:1..221,AF109119.1:1..234,1..2033)/genes="HRNBPI"
mRNA	join(AF109107.1:372..711,AF109108.1:173..415,AF109109.1:166..309,AF109110.1:94..147,AF109111.1:75..167,AF109112.1:50..110,AF109113.1:200..253,AF109114.1:22..102,AF109115.1:67..199,AF109116.1:108..150,AF109117.1:35..99,AF109118.1:134..176,AF109119.1:62..137,159..2033)/genes="HRNBPI"
	/product="hexaribonucleotide binding protein 1 isoform beta"
mRNA	/notes="alternatively spliced"
	join(AF109107.1:372..711,AF109108.1:173..415,AF109109.1:166..309,AF109110.1:94..147,AF109111.1:75..167,AF109112.1:50..110,AF109113.1:200..253,AF109114.1:22..102,AF109115.1:67..199,AF109116.1:108..150,AF109117.1:35..99,AF109119.1:62..137,159..2033)/genes="HRNBPI"
	/product="hexaribonucleotide binding protein 1 isoform gamma"
mRNA	join(AF109107.1:372..711,AF109108.1:173..415,AF109109.1:166..309,AF109110.1:94..147,AF109111.1:75..167,AF109112.1:50..110,AF109113.1:200..253,AF109114.1:22..102,AF109115.1:67..199,AF109116.1:108..150,AF109117.1:35..99,AF109119.1:62..137,237..2033)/genes="HRNBPI"
	/product="hexaribonucleotide binding protein 1 isoform alpha"
CDS	/notes="alternatively spliced"
	join(AF109107.1:625..711,AF109108.1:173..415,AF109109.1:166..309,AF109110.1:94..147,AF109111.1:75..167,AF109112.1:50..110,AF109113.1:200..253,AF109114.1:22..102,AF109115.1:67..199,AF109116.1:108..150,AF109117.1:35..99,AF109119.1:62..137,159..281)/genes="HRNBPI"
	/codon_start=1
	/product="hexaribonucleotide binding protein 1 isoform gamma"

```

gamma "
/db_xref="GI:1903238"
/translation="MLASQGVLLHPYGVPMIVPAAPYLPGLIQNQEAAPADTMAOP
YASQAPPPONGIPAEYTAHPHPAPYTGTTVPEHTLNLYPPAOTHSQSPADTSA
QTVSGTATOTDDAAPTQGPOTQSENTENKSPKRLHVSNI PRFRDPDLRQMFQGF
KILIDVELIFNERGSKGFVTFENSADADAREKLHGTVEGRKIEVNATARVMTN
KKTVPITNGWKLNVVGVGAYSPFYAGTVLLCOANQEGSSMSAPSLVTSAMPGF
PYPAATAAAYRGHAHLRGRTVYNTFRAAAPPPPIPAYGGVVYQEPVYGNKLLQGGY
AAYRYAQPTATAAAYSDSYGRVVAADPYHHALAPATPYGVGMANAPFLTDKTRSH
ADDVGVLLSSLOASIRGGYNRFPAPY"
join(AFI09107.1:625..711,AFI09108.1:173..415,
AFI09109.1:166..309,AFI09110.1:94..147,AFI09111.1:75..167,
AFI09112.1:50..110,AFI09113.1:200..253,AFI09114.1:22..102,
AFI09115.1:67..199,AFI09116.1:108..150,AFI09117.1:35..99,
AFI09119.1:62..137,237..281)
/gene="HRNBPI"
/notes="similar to Caenorhabditis elegans fox-1;
alternatively spliced"
/codon_start=1
/product="hexaribonucleotide binding protein 1 isoform
alpha"
/protein_id="AAL83408.1"
/db_xref="GI:1903238"
/translation="MLASQGVLLHPYGVPMIVPAAPYLPGLIQNQEAAPADTMAOP
YASQAPPPONGIPAEYTAHPHPAPYTGTTVPEHTLNLYPPAOTHSQSPADTSA
QTVSGTATOTDDAAPTQGPOTQSENTENKSPKRLHVSNI PRFRDPDLRQMFQGF
KILIDVELIFNERGSKGFVTFENSADADAREKLHGTVEGRKIEVNATARVMTN
KKTVPITNGWKLNVVGVGAYSPFYAGTVLLCOANQEGSSMSAPSLVTSAMPGF
PYPAATAAAYRGHAHLRGRTVYNTFRAAAPPPPIPAYGGVVYQEPVYGNKLLQGGY
AAYRYAQPTATAAAYSDSYGRVVAADPYHHALAPATPYGVGMANASYRGGYNRFPAPY
"
join(AFI09107.1:625..711,AFI09108.1:173..415,
AFI09109.1:166..309,AFI09110.1:94..147,AFI09111.1:75..167,
AFI09112.1:50..110,AFI09113.1:200..253,AFI09114.1:22..102,
AFI09115.1:67..199,AFI09116.1:108..150,AFI09117.1:35..99,
AFI09118.1:124..176,AFI09119.1:62..137,159)
/gene="HRNBPI"
/notes="similar to Caenorhabditis elegans fox-1;
alternatively spliced"
/codon_start=1
/product="hexaribonucleotide binding protein 1 isoform
beta"
/protein_id="AAL83407.1"
/db_xref="GI:19032384"
/translation="MLASQGVLLHPYGVPMIVPAAPYLPGLIQNQEAAPADTMAOP
YASQAPPPONGIPAEYTAHPHPAPYTGTTVPEHTLNLYPPAOTHSQSPADTSA
QTVSGTATOTDDAAPTQGPOTQSENTENKSPKRLHVSNI PRFRDPDLRQMFQGF
KILIDVELIFNERGSKGFVTFENSADADAREKLHGTVEGRKIEVNATARVMTN
KKTVPITNGWKLNVVGVGAYSPFYAGTVLLCOANQEGSSMSAPSLVTSAMPGF
PYPAATAAAYRGHAHLRGRTVYNTFRAAAPPPPIPAYGGVVYQEPVYGNKLLQGGY
AAYRYAQPTATAAAYSDRNQFVFVADEISCHTSVATDSEMLPTPTTHLLQPPPTA
LVPT
159..2033
/gene="HRNBPI"
/notes="alternate exon 14"
/number=14
237..2033
/gene="HRNBPI"
/notes="alternate exon 14"
/number=14
2017..2022
/gene="HRNBPI"
2033
polyA_signal
polyA_site
COUNT 922 a 521 c 586 g 942 t
IN
5.2%; Score 70; DB 9; Length 2971;
ct Local Similarity 100.0%; Pred.No.1.3e-30;
ches 70; Conservaive 0; Mismatches 0; Indels 0; Gaps 0;
1045 GATGCCAAGACTAGGCGCATCGTATGATGGGTCTCGTTCTTCTTCATTGCAGCCT 1104

```

```
Query Match          5.2%; Score 70; DB 9; Length 2971;
Best Local Similarity 100.0%; Pred. No. 1.3e-30;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 1045 GATGCCAAGACTAGGAGCCATGCTGATGATGTGGGCTCTCGTTCTTCTTCA TTGCAGGCT 1104

```
TITLE Direct Submission
JOURNAL Submitted (28-JAN-2000) Internal Medicine/Hematology-Oncology,
University of Cincinnati College of Medicine, The Vontz Center for
Molecular Studies, 3125 Eden Avenue, Cincinnati, OH 45267-0508, USA
FEATURES
  source      Location/Qualifiers
    1..2000
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="16"
      /map="16p13.3"
    1..2000
      /genes="HRNBPI"
      /gene="HRNBPI"
      /note="RNA-binding protein"
      /codon_start=1
      /product="hexaribonucleotide binding protein 1 isoform
      gamma"
      /protein_id="AAU71904.1"
      /db_xref="GI:18461367"
      /translation="MLASQGVLLHPYGVPMVPAAPYLPGLIQGNEAAAAADPTWAQP
      YASAOFPAPONGPAEYTPHPHPAPEYTGOTTVPHEITNLNLYPPAOTHSSEQSPADTSA
      OTVSGTATDDAAPTGGOPQTOPSENTENKSKOPKELHVSNIPIFRDPLDRMFQGF
      GKILDVEIIFNERSKGFVTEENSADADREKELHGTVEGRKIEVNATRVMTN
      KKTNYITNGKLNPNVGVAVSPFYAGTVLLCOAQEGSSMTSAPSLVYTSAMEGF
      PYPAATAAAYRGALHGRGRVTYNTFRAAAPPPPIPAYGGVYQEPVYGNKLLGGY
      AAYRYAQPATATAAAYSDSYGRVYAADPYHHALAPATPYGVGAMNAPFLTDKTRSH
      ADDGVLVLSLQASIYRGGRFAPY"
    502 a 505 c 487 g 506 t
BASE COUNT
ORIGIN
  Query Match 5.2%; Score 70; DB 9; Length 2000;
  Best Local Similarity 100.0%; Pred. No. 1.3e-30;
  Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1045 GATGCCAAGACTAGGAGCCATGCTGATGATGTCGGTCTCGTCTTCTTCTTCATTGCAGGCT 1104
Db 1411 GATGCCAAGACTAGGAGCCATGCTGATGATGTCGGTCTCGTCTTCTTCTTCATTGCAGGCT 1470
Qy 1105 AGTATATACC 1114
Db 1471 AGTATATACC 1480
RESULT 10
AF107203
LOCUS Homo sapiens ataxin 2-binding protein (A2BP) mRNA linear PRI 23-JUN-2000
DEFINITION Homo sapiens ataxin 2-binding protein (A2BP) mRNA, complete cds.
ACCESSION AF107203
VERSION AF107203.1 GI:8671585
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE
  1 (bases 1 to 2279)
  Shibata,H., Huynh,D.P., Vo,T.T. and Pulst,S.-M.
  A novel protein, A2BP, with RNA binding motif binds to C-terminal
  ataxin-2
  Unpublished
  2 (bases 1 to 2279)
  Shibata,H., Huynh,D.P., Vo,T.T. and Pulst,S.-M.
  Direct Submission
  Submitted (16-NOV-1998) Division of Neurology and Neurogenetics
  Laboratory, Burns and Allen Research Institute, Cedars-Sinai
  Medical Center, UCLA School of Medicine, 8700 Beverly Blvd., Los
  Angeles, CA 90048, USA
  Location/Qualifiers
    1..2279
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /note="similar to Homo sapiens cDNA clone HSP38A20"
    1..2279
gene
```

```
CDS
  /gene="A2BP"
  1047..2180
  /gene="A2BP"
  /codon_start=1
  /product="ataxin 2-binding protein"
  /protein_id="AAF78291.1"
  /db_xref="GI:8671586"
  /translation="MAQPYASAOFPAPONGIPAEYTAHPHPAPEYTGOTTVPHEITLN
  LYPPAOTHSSEQSPADTSAQTVSGTATQDDAAAPTGGOPQTOPQSENTENKSKPRLHVS
  NIPFRDPLDRMFQGFQKILDVEIIFNERSKGFVTEENSADADREKELHGTV
  VEGRKIEVNATRVMTNPNVNGKLNPNVGVAVSPFYAGTVLLCOAQEGS
  SMYSAPSLVYTSAMEGFYPAAATAAAYRGALHGRGRVTYNTFRAAAPPPPIPAYG
  GVYQDGFYADLYGVAARYAQPATATAAAYSDSYGRVYAADPYHHALAPATPYGV
  GAMNAPFLTDKTRSHADDGVLVLSLQASIYRGGRFAPY"
  1341..1358
misc_feature
  /gene="A2BP"
  /note="Region: RNA binding motif RNP-2"
  1452..1472
misc_feature
  /gene="A2BP"
  /note="Region: RNA binding motif RNP-1"
  497 a 712 c 636 g 434 t
BASE COUNT
ORIGIN
  Query Match 5.2%; Score 70; DB 9; Length 2279;
  Best Local Similarity 100.0%; Pred. No. 1.3e-30;
  Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1045 GATGCCAAGACTAGGAGCCATGCTGATGATGTCGGTCTCGTCTTCTTCTTCATTGCAGGCT 1104
Db 2079 GATGCCAAGACTAGGAGCCATGCTGATGATGTCGGTCTCGTCTTCTTCTTCATTGCAGGCT 2138
Qy 1105 AGTATATACC 1114
Db 2139 AGTATATACC 2148
RESULT 11
AR134676
LOCUS Sequence 1 from patent US 6194171.
DEFINITION AR134676
ACCESSION AR134676
VERSION AR134676.1 GI:14123581
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE
  1 (bases 1 to 2372)
  Pulst,S.M. and Shibata,H.
  Nucleic acids encoding ataxin-2 binding proteins
  Patent: US 6194171-A 1 27-FEB-2001,
  Location/Qualifiers
    1..2372
      /organism="unknown"
      /organism="unknown"
    548 a 726 c 646 g 452 t
BASE COUNT
ORIGIN
  Query Match 5.2%; Score 70; DB 6; Length 2372;
  Best Local Similarity 100.0%; Pred. No. 1.3e-30;
  Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1045 GATGCCAAGACTAGGAGCCATGCTGATGATGTCGGTCTCGTCTTCTTCTTCATTGCAGGCT 1104
Db 2133 GATGCCAAGACTAGGAGCCATGCTGATGATGTCGGTCTCGTCTTCTTCTTCATTGCAGGCT 2192
Qy 1105 AGTATATACC 1114
Db 2193 AGTATATACC 2202
RESULT 12
HSFOX14
LOCUS Homo sapiens hexaribonucleotide binding protein 1 alpha, beta, and
DEFINITION
```

Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
(E-mail: khashi@nih.go.jp, URL: http://www.nih.go.jp/yoken/genebank/,
Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)
Lab host: TOP10
Vector: pME18S-FL3 (Acc.No. AB009864)
R. Site1: DraIII (CACCATGTG)
R. Site2: DraIII (CACCATGTG)
Description: 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCTTTTTTTTTTTTT]; double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18S-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by oligo-capping method
(Sugano et al., Institute of Medical Science, University of Tokyo).

Custom primer used for sequencing
(5' end primer [CTTCTGCTCTAAAGCTGCG];
3' end primer [CGACCTGCAGCTCGACACA]).

FEATURES

source

Location/Qualifiers

1..1475

/organism="Macaca fascicularis"

/db_xref="taxon:9541"

/clone="QtrA-11594"

/sex="male"

/tissue_type="temporal lobe right"

/clone_lib="macaque brain cDNA library QtrA"

/dev_stage="adult"

90..1220

CDS

/codon_start=1

/product="hypothetical protein"

/protein_id="BAB46877.1"

/db_xref="GI:13874511"

/translation="MEEGKSRNVQGNQAEAAAPDTMAQPYASAFAPFQNGIPAEYTA
APHPAPEYTGQTTVPHTLNLYPPAQTHSEQSPADTSAQTSVGTATQDDAAPTGG
QPQTSPENSTENKSPKRLHVSNIPIFRFPDRLRMFGQFGKILDEIIFNERSKGF
GVTENSADADAREKLHGTVEGRKIEVNNTARVMTNKTVPYTNWKLNPVVG
AVYSPFYAGTVLLCAQNEGSSMYSPSSLVYTSAMPGFYPAATAAAYRGALRG
RGRTVYNTFRAAAPPPPIPAYGGVYQDGFYGGADYGGYAAVRYAQPTATAAAYSDR
NQVFVADEISCSNTSAVTDFEMLPTPTITLLQPPPTALVP"

389 a 414 c 363 g 309 t

BASE COUNT
ORIGIN

Query Match 5.2%; Score 70; DB 9; Length 1475;
Best Local Similarity 100.0%; Pred. No. 1.3e-30;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1045 GATGCCAAGACTAGGAGCCATGCTGATGTGGGTCTCGTCTCTTTCTTCATTGCAGGCT 1104

|||||

Db 1241 GATGCCAAGACTAGGAGCCATGCTGATGTGGGTCTCGTCTCTTTCTTCATTGCAGGCT 1300

|||||

QY 1105 AGTATATACC 1114

|||||

Db 1301 AGTATATACC 1310

RESULT 8
AK001027

Homo sapiens cDNA FLJ10165 fib, clone HEMBA1003591, weakly similar
to CHLOROPLAST 28 KD RIBONUCLEOPROTEIN PRECURSOR.

AK001027 1513 bp mRNA linear PRI 01-AUG-2002

DEFINITION

oligo capping; fib (full insert sequence).

ACCESSION

AK001027.1 GI:7022045

VERSION

oligo capping; fib (full insert sequence).

KEYWORDS

mRNA, clone lib:HEMBA1 clone:HEMBA1003591..

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

1

REFERENCE

Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,

1

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Nishikawa,T., Negai,K., Sugano,S., Shiratori,A., Sudo,H.,
Wagatsuma,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
Nakamura,Y., Nagahara,K., Masuno,Y., Ninomiya,K. and Iwayanagi,T.
NEDO human cDNA sequencing project
UNPUBLISHED
2 (bases 1 to 1513)
Isogai,T. and Otsuki,T.
Direct Submission
Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5' - & 3' -end one pass sequencing and clone selection;
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.

FEATURES

source

Location/Qualifiers

1..1513

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="HEMBA1003591"

/tissue_type="whole embryo, mainly head"

/clone_lib="HEMBA1"

/dev_stage="embryo, 10 weeks"

/note="cloning vector: pME18SFL3"

221..1414

/note="unnamed protein product"

/codon_start=1

/protein_id="BAA91472.1"

/db_xref="GI:7022046"

/translation="MNCEREQLRQNEQAEAAAPDTMAQPYASAFAPFQNGIPAEYTA
PHPAPEYTGQTTVPHTLNLYPPAQTHSEQSPADTSAQTSVGTATQDDAAPTGG
QPQTSPENSTENKSPKRLHVSNIPIFRFPDRLRMFGQFGKILDEIIFNERSKGF
GVTENSADADAREKLHGTVEGRKIEVNNTARVMTNKTVPYTNWKLNPVVG
AVYSPFYAGTVLLCAQNEGSSMYSPSSLVYTSAMPGFYPAATAAAYRGALRG
RGRTVYNTFRAAAPPPPIPAYGGVYQDGFYGGADYGGYAAVRYAQPTATAAAYSDY
RYAADPYHHALAPAPTYGVGMNAPLTDKTRSHADDVGLVLSLQASLIYRGY
RFAPY"

BASE COUNT 354 a 418 c 395 g 346 t

ORIGIN

Query Match 5.2%; Score 70; DB 9; Length 1513;
Best Local Similarity 100.0%; Pred. No. 1.3e-30;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1045 GATGCCAAGACTAGGAGCCATGCTGATGTGGGTCTCGTCTCTTTCTTCATTGCAGGCT 1104

|||||

Db 1313 GATGCCAAGACTAGGAGCCATGCTGATGTGGGTCTCGTCTCTTTCTTCATTGCAGGCT 1372

|||||

QY 1105 AGTATATACC 1114

|||||

Db 1373 AGTATATACC 1382

RESULT 9
AF229057

LOCUS

DEFINITION

Homo sapiens hexaribonucleotide binding protein 1 isoform gamma

(HRNB1) mRNA, complete cds.

ACCESSION

AF229057.1 GI:18461366

VERSION

AF229057.1

KEYWORDS

complete cds.

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 2000)

REFERENCE

Chen,W. and Winkelmann,J.C.

1

AUTHORS

AF229057 2000 bp mRNA linear PRI 01-FEB-2002
Homo sapiens hexaribonucleotide binding protein 1 isoform gamma
(HRNB1) mRNA, complete cds.

AF229057.1 GI:18461366

DEFINITION

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 2000)

REFERENCE

Chen,W. and Winkelmann,J.C.

1

AUTHORS

Consensus quality: 23573 bases at least Q40
 Consensus quality: 25350 bases at least Q30
 Consensus quality: 26823 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 44 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

1 1065: contig of 1065 bp in length
 * 1066: gap of unknown length
 * 1166: contig of 1246 bp in length
 * 2112: gap of unknown length
 * 2511: gap of unknown length
 * 2512: contig of 1471 bp in length
 * 3982: gap of unknown length
 * 4082: contig of 1514 bp in length
 * 4083: gap of unknown length
 * 5596: gap of unknown length
 * 5597: contig of 1385 bp in length
 * 7081: gap of unknown length
 * 7082: gap of unknown length
 * 7181: gap of unknown length
 * 8496: contig of 1315 bp in length
 * 8596: gap of unknown length
 * 8497: contig of 1504 bp in length
 * 10100: contig of 1504 bp in length
 * 10101: gap of unknown length
 * 10200: gap of unknown length
 * 10201: contig of 1049 bp in length
 * 11349: gap of unknown length
 * 11350: contig of 1352 bp in length
 * 12701: gap of unknown length
 * 12801: contig of 1411 bp in length
 * 12802: gap of unknown length
 * 14212: contig of 1411 bp in length
 * 14213: gap of unknown length
 * 14313: contig of 1068 bp in length
 * 15380: gap of unknown length
 * 15381: contig of 1036 bp in length
 * 16516: gap of unknown length
 * 16517: gap of unknown length
 * 17946: contig of 1330 bp in length
 * 17947: gap of unknown length
 * 18046: gap of unknown length
 * 18047: contig of 1100 bp in length
 * 19146: gap of unknown length
 * 19147: contig of 1280 bp in length
 * 20526: gap of unknown length
 * 20527: contig of 1104 bp in length
 * 20627: gap of unknown length
 * 21730: contig of 1104 bp in length
 * 21731: gap of unknown length
 * 21830: contig of 1528 bp in length
 * 23358: gap of unknown length
 * 23359: contig of 1880 bp in length
 * 23459: gap of unknown length
 * 25338: contig of 1200 bp in length
 * 25438: gap of unknown length
 * 26639: contig of 1630 bp in length
 * 26739: gap of unknown length
 * 28369: contig of 1428 bp in length
 * 28469: gap of unknown length
 * 29897: gap of unknown length
 * 29997: contig of 1054 bp in length
 * 31050: gap of unknown length
 * 31150: gap of unknown length
 * 31551: contig of 1435 bp in length
 * 32585: gap of unknown length
 * 32686: contig of 1407 bp in length
 * 34092: gap of unknown length
 * 34192: contig of 1299 bp in length
 * 34993: gap of unknown length
 * 35492: contig of 1505 bp in length
 * 35592: gap of unknown length
 * 37097: contig of 1116 bp in length
 * 38312: gap of unknown length
 * 38313: contig of 2167 bp in length
 * 38413: gap of unknown length
 * 40579: contig of 2032 bp in length
 * 40580: gap of unknown length
 * 40680: contig of 2032 bp in length
 * 42711: gap of unknown length
 * 42712: contig of 1413 bp in length
 * 44224: contig of 1413 bp in length

* 44225: gap of unknown length
 * 44324: contig of 1535 bp in length
 * 45859: gap of unknown length
 * 45959: contig of 2016 bp in length
 * 45960: gap of unknown length
 * 47975: contig of 1974 bp in length
 * 48076: gap of unknown length
 * 50149: gap of unknown length
 * 50050: contig of 1551 bp in length
 * 50150: gap of unknown length
 * 51701: contig of 1751 bp in length
 * 51800: gap of unknown length
 * 53551: gap of unknown length
 * 53552: contig of 2476 bp in length
 * 53652: gap of unknown length
 * 56128: contig of 1693 bp in length
 * 56227: gap of unknown length
 * 58228: contig of 2237 bp in length
 * 59221: gap of unknown length
 * 60257: contig of 2481 bp in length
 * 60258: gap of unknown length
 * 62358: contig of 1901 bp in length
 * 62359: gap of unknown length
 * 62359: contig of 1938 bp in length
 * 64840: gap of unknown length
 * 64940: contig of 2879 bp in length
 * 66878: gap of unknown length
 * 66977: gap of unknown length
 * 69566: contig of 1760 bp in length
 * 69567: gap of unknown length
 * 71716: contig of 2113 bp in length
 * 71717: gap of unknown length
 * 71816: contig of 2113 bp in length
 * 73929: contig of 2113 bp in length
 * 71917

FEATURES

source

1..73929
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-33J14"

BASE COUNT 19038 a 16649 c 14558 g 19363 t 4321 others
 ORIGIN

Query Match 5.9% Score 79; DB 2; Length 73929;

Best Local Similarity 100.0%; Pred. No. 4.8e-36; Indels 0; Gaps 0;
 Matches 79; Conservative 0; Mismatches 0;

QY 945 CAGTTACGACGAGTTATGTCGCCACCCCTACACACACACTTCTCCAGCCCCAC 1004

Db 11590 CAGTTACGACGAGTTATGTCGCCACCCCTACACACACACTTCTCCAGCCCCAC 11649

QY 1005 CTACGGCGTTGGTGCCATG 1023

Db 11650 CTACGGCGTTGGTGCCATG 11668

RESULT 7

AB060859

LOCUS 1475 bp mRNA linear PRI 13-JUN-2001
 DEFINITION Macaca fascicularis brain cDNA clone:QtrA-11594, full insert

ACCESSION

AB060859.1 GI:13874510

VERSION oligo capping; fis (full insert sequence).

KEYWORDS Macaca fascicularis adult male temporal lobe right cDNA to mRNA,

SOURCE clone lib:macaque brain cDNA library QtrA clone:QtrA-11594.

ORGANISM

Macaca fascicularis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 Cercopithecinae; Macaca.

REFERENCE

AUTHORS

Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M.,

Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K.

TITLE Isolation of full-length cDNA clones from macaque brain cDNA

libraries

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1475)

AUTHORS Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.

TITLE Direct Submission

JOURNAL Submitted (27-APR-2001) Katsuyuki Hashimoto, National Institute of

QY 1263 GACATCACTTTAGCACTCAAGAAACAAC 1292
|||||
Db 112315 GACATCACTTTAGCACTCAAGAAACAAC 112286
|||||

RESULT 5
AF191501
LOCUS
DEFINITION Mus musculus hexaribonucleotide binding protein 1 (Hrnbp1) mRNA,
complete cds.
ACCESSION AF191501
VERSION AF191501.1 GI:19032413
KEYWORDS
SOURCE
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1586)
Chen, W. and Winkelman, J.C.
Submitted (01-OCT-1999) Internal Medicine/Hematology-Oncology,
University of Cincinnati College of Medicine, 231 Bethesda Ave.,
Cincinnati, OH 45267-0508, USA
CINCINNATI, OH 45267-0508, USA
Location/Qualifiers
1..1586
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="skeletal muscle"
1..1586
/gene="Hrnbp1"
166..1419
/gene="Hrnbp1"
/codon_start=1
/product="hexaribonucleotide binding protein 1"
/protein_id="AAL83425.1"
/db_xref="GI:19032414"

gene
CDS

BASE COUNT 407 a 458 c 374 g 347 t
ORIGIN

Query Match 12.0%; Score 161; DB 10; Length 1586;
Best Local Similarity 100.0%; Pred. No. 3.9e-86;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 556 ACCCTTACCAATGGCTGGAATTAATCCAGTTGTGGCGCGGTCTACAGCCCGAC 615
|||||
Db 829 ACCCTTACCAATGGCTGGAATTAATCCAGTTGTGGCGCGGTCTACAGCCCGAC 888
|||||

QY 616 TTCTATGACGACGGCTGCTGTGTGCCAGGCCAACACGAGGATCTTCCATGTACAGT 675
|||||
Db 889 TTCTATGACGACGGCTGCTGTGTGCCAGGCCAACACGAGGATCTTCCATGTACAGT 948
|||||

QY 676 GGGCCAGTTCATCTGTATATATCTTCTGCAATGCCCTGGCTT 716
|||||
Db 949 GGGCCAGTTCATCTGTATATATCTTCTGCAATGCCCTGGCTT 989
|||||

RESULT 6
AC103207
LOCUS
DEFINITION Rattus norvegicus clone CH230-33J14, *** SEQUENCING IN PROGRESS
***, 44 unordered pieces.
ACCESSION AC103207
VERSION AC103207.3 GI:21731168
KEYWORDS HTG; HTGS_PHASE1.

SOURCE ORGANISM Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 73929)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
Albrooks, S.L., Amarantunga, H.C., Are, J.R., Ayale, M., Banks, T.,
Barbata, J., Benton, J., Bimaga, K., Blankenburg, K., Bryant, D.,
Buck, J., Bowie, S., Brieve, M., Brown, M., Brown, M., Brown, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,
Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karleson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louie, H.,
Lorado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okuwonu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.B., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Uman, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Zhou, J., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 73929)
Worley, K.C.
Direct Submission
Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 73929)
Worley, K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:17974697.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GJLU
Center clone name: CH230-33J14
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990129

Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
 Burch, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Haviak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
 Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Lucier, R., Luna, R., Ma, J.,
 Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenwo, S., Oguh, M., Okwodu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,
 Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, K., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Teifrod, B., Thomas, N., Thomas, S.,
 Ugami, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstock, G., and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 115860)
 Worley, K.C.

Direct Submission
 Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 115860)
 Worley, K.C.

Direct Submission
 Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 18, 2002 this sequence version replaced gi:20514522.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GXLG
 Center clone name: CH230-34C2
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 99693 bases at least Q40
 Consensus quality: 101936 bases at least Q30
 Consensus quality: 102743 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length.
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 21 contigs. The true order of the pieces
 * is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1347: contig of 1347 bp in length
 * 1348 1447: gap of unknown length
 * 1448 2834: contig of 1387 bp in length
 * 2835 2934: gap of unknown length
 * 2935 4253: contig of 1319 bp in length
 * 4254 4353: gap of unknown length
 * 4354 5477: contig of 1124 bp in length
 * 5478 5577: gap of unknown length
 * 5578 7584: contig of 1907 bp in length
 * 7585 9406: contig of 1822 bp in length
 * 9407 9506: gap of unknown length
 * 9507 11712: contig of 2206 bp in length
 * 11713 11812: gap of unknown length
 * 11813 13462: contig of 1650 bp in length
 * 13463 13562: gap of unknown length
 * 13563 17222: contig of 3660 bp in length
 * 17223 17322: gap of unknown length
 * 17323 20617: contig of 3295 bp in length
 * 20618 20717: gap of unknown length
 * 20718 24825: contig of 4108 bp in length
 * 24826 24925: gap of unknown length
 * 24926 28811: contig of 3886 bp in length
 * 28812 28911: gap of unknown length
 * 28912 33686: contig of 4775 bp in length
 * 33687 33786: gap of unknown length
 * 33787 40317: contig of 6431 bp in length
 * 40318 40317: gap of unknown length
 * 40319 48012: contig of 7595 bp in length
 * 48013 56482: contig of 8470 bp in length
 * 56483 56582: gap of unknown length
 * 56583 66204: contig of 9622 bp in length
 * 66205 74546: contig of 8242 bp in length
 * 74547 74546: gap of unknown length
 * 74548 86522: contig of 11876 bp in length
 * 86523 86622: gap of unknown length
 * 86623 100049: contig of 13427 bp in length
 * 100050 100149: gap of unknown length
 * 100150 115860: contig of 15711 bp in length.

FEATURES

Location/Qualifiers
 1..115860
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-34C2"

BASE COUNT 33039 a 22417 c 22786 g 30574 t 7044 others
 ORIGIN

Query Match 12.5%; Score 168; DB 2; Length 115860;
 Best Local Similarity 99.3%; Pred. No. 2.5e-90;
 Matches 268; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1023 GAATGCTTTTGGCCCTTACCGGATGCCAGACTAGGAGCCATGCTGATGATGTTGGGTCT 1082

Db 112555 GAATGCTTTTGGCCCTTACCGGATGCCAGACTAGGAGCCATGCTGATGATGTTGGGTCT 112496

Qy 1083 CGTCTTTTCTTCATTCAGGCTAGTATATACCAAGGGGATACAAACCGTTTGTCTCCATA 1142

Db 112495 CGTCTTTTCTTCATTCAGGCTAGTATATACCAAGGGGATACAAACCGTTTGTCTCCATA 112436

Qy 1143 TTAATGATAAACCATTAAACAACAAGCAAAAAACAACAAAAACAAAAACCAAC 1202

Db 112435 TTAATGATAAACCATTAAACAACAAGCAAAAAACAACAAAAACAAAAACCAAC 112376

Qy 1203 CTTCCAAATGTGGGAG 1262

Db 112375 CTTCCAAATGTGGGAG 112316

GGVYODGFGADYGGYAAAYRQAQTPATAAAYSDSYGRVYAADPYHHTLAPATYTG
VGMNAPALTDKTRSHADDVGLVLSLQASIRYGGYNRFAPY"

misc_feature 358..375
/gene="A2bp"
/notes="Region: RNA binding motif RNP-2"
misc_feature 475..495
/gene="A2bp"
/notes="Region: RNA binding motif RNP-1"
BASE COUNT 362 a 396 c 324 g 281 t
ORIGIN

Query Match 14.5%; Score 194; DB 10; Length 1363;
Best Local Similarity 99.6%; Pred. No. 2.9e-106;
Matches 244; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 556 AACCCCTACACCAATGCTGGAATTAATCCAGTTGTCGGCGGCTCTACAGCCCGGAC 615
DB 616 AACCCCTACACCAATGCTGGAATTAATCCAGTTGTCGGCGGCTCTACAGCCCGGAC 675
QY 616 TTCTATGACGACGGTGTGTTGTGTCAGGCGCAACAGGAGGATCTTCCATGTACAGT 675
DB 676 TTCTATGACGACGGTGTGTTGTGTCAGGCGCAACAGGAGGATCTTCCATGTACAGT 735
QY 676 GGGCCAGTTCACTTGTATATCTTTCGCAATGCTGCTTTCATATCCGGCGGCACT 735
DB 736 GGGCCAGTTCACTTGTATATCTTTCGCAATGCTGCTTTCATATCCGGCGGCACT 795
QY 736 GCTGCAGTGCATACCGAGGGGTCACCTTCGAGGCGGTCGTCAGCGGTGTACACACC 795
DB 796 GCTGCAGTGCATACCGAGGGGTCACCTTCGAGGCGGTCGTCAGCGGTGTACACACC 855
QY 796 TTCAG 800
DB 856 TTCAG 860

RESULT 3
AB041596
LOCUS
DEFINITION
AB041596
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
COMMENT

AB041596 2002 bp mRNA linear ROD 30-JUN-2000
Mus musculus brain cDNA, clone MNCB-3035, similar to Homo sapiens
CDNA FLJ10165, clone HEMBA1003591.
AB041596
1 (sites)
fis (full insert sequence).
Mus musculus (strain: C57BL) adult female cDNA to mRNA,
clone lib:Sugano mouse brain mncb clone:MNCB-3035.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (sites)
Osada, N., Kusuda, J., Tanuma, R., Ito, A., Hirata, M., Sugano, S. and
Hashimoto, K.
isolation of full-length cDNA clones from mouse brain cDNA library
made by oligo-capping method
Unpublished
2 (bases 1 to 2002)
Hashimoto, K., Osada, N., Kusuda, J. and Sugano, S.
Direct Submission
Submitted (12-APR-2000) Katsuyuki Hashimoto, National Institute of
Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
(E-mail: khashi@nih.go.jp, URL: http://www.nih.go.jp/yoken/genebank/,
Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)
URL: http://www.nih.go.jp/yoken/genebank/
Lib Name: Sugano mouse brain mncb
Lab host: TOPI0
Vector: pME18S-FL3
1st strand cDNA was primed with an oligo(dT) primer
[ATGCGCTTTTCTTTTCTTTTCTTTT]; double-stranded cDNA was synthesized
using specific 5' and 3' primers and amplified by PCR. The PCR
product was digested with SfiI and size selection was performed to
exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
into distinct DraIII sites of pME18S-FL3. XhoI sites just outside

the DraIII sites can be used to isolate the cDNA insert. Library
was constructed by Sugano et al. (University of Tokyo, Institute of
Medical Science). Custom primer used for sequencing (5' end primer
(CTTCTGCTTAAAGCTGG); 3' end primer
(CGACCTGACGCTCGAGCACA).
A part of this sequence is reported in AU067167.

FEATURES
source
1..2002
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="MNCB-3035"
/sex="female"
/clone_lib="Sugano mouse brain mncb"
/dev_stage="adult"
637..1827
/note="unnamed protein product"
/codon_start=1
/protein_id="BAA95079.1"
/db_xref="GI:7670456"
/translation="MNCERQLRQNEAAADPTMAQPYASAFQAPPQNGIPAEYTPAP
HPHPAETVGTTPVHTLNLYPPTQTHSEASADTSATVSGTATATDAAPTDGOPO
TOPSENTESKQPKLHVSNIPFRDPDLROMFGQFGKILDEIIFNERSKGGFV
TFENADADRAKREKJHGVVVEGRKIEVNNATRVNKTNPYINGWKNPVGAVY
SPDFYAGTVLLICQANQEGSSMYSGPSLVYTSAMPFPYPAATAAAAYGAHLRGRV
TVYNTFRAAAPPPPIPAYGGVYODGYGADYGGYAAAYRQAQTPATAAAYSDSYGR
VYAADPYHHTLAPATYTGGAATTAATCCAGTTGTCGGCGGCTCTACAGCCCGGAC
FAPY"
BASE COUNT 518 a 589 c 516 g 379 t
ORIGIN

Query Match 14.5%; Score 194; DB 10; Length 2002;
Best Local Similarity 99.6%; Pred. No. 2.9e-106;
Matches 244; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 556 AACCCCTACACCAATGCTGGAATTAATCCAGTTGTCGGCGGCTCTACAGCCCGGAC 615
DB 1240 AACCCCTACACCAATGCTGGAATTAATCCAGTTGTCGGCGGCTCTACAGCCCGGAC 1299
QY 616 TTCTATGACGACGGTGTGTTGTGTCAGGCGCAACAGGAGGATCTTCCATGTACAGT 675
DB 1300 TTCTATGACGACGGTGTGTTGTGTCAGGCGCAACAGGAGGATCTTCCATGTACAGT 1359
QY 676 GGGCCAGTTCACTTGTATATCTTCTGCAATGCTGCTTTCATATCCGGCGGCACT 735
DB 1360 GGGCCAGTTCACTTGTATATCTTCTGCAATGCTGCTTTCATATCCGGCGGCACT 1419
QY 736 GCTGCAGTGCATACCGAGGGGCTCACTTCGAGGCGGTCGTCGACCGGTGTACACACC 795
DB 1420 GCTGCAGTGCATACCGAGGGGCTCACTTCGAGGCGGTCGTCGACCGGTGTACACACC 1479
QY 796 TTCAG 800
DB 1480 TTCAG 1484

RESULT 4
AC120661/c
LOCUS
DEFINITION
AC120661
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

AC120661 115860 bp DNA linear HTG 23-JUL-2002
Rattus norvegicus clone CH230-34C2, *** SEQUENCING IN PROGRESS ***,
21 unordered pieces.
AC120661
AC120661.2 GI:21902877
HTG: HTGS_PHASE1.
Rattus norvegicus
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus
1 (bases 1 to 115860)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-oshman, F.R., Allen, C.,
Albrooks, S.L., Amaral, H.C., Are, J.R., Ayele, M., Banks, T.,
Barbaria, J., Benton, J., Bimane, K., Blankenburg, K., Bonnin, D.,

JOURNAL Patent: WO 0174901-A 1 11-OCT-2001;
Scios Inc. (US)
FEATURES Location/Qualifiers
source 1..1340
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
BASE COUNT 366 a 394 c 317 g 263 t
ORIGIN

Query Match 100.0%; Score 1340; DB 6; Length 1340;
Best local Similarity 100.0%; Pred. No. 0;
Matches 1340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGGCCGCGCTGACACAAATGGCTCAGTTATGCTCAGCGAGTTGCTCCACCCGAGA 60
Db 1 GCGGCCGCGCTGACACAAATGGCTCAGTTATGCTCAGCGAGTTGCTCCACCCGAGA 60

Qy 61 ATGCGATCCTGAGATACAGCGCCCTCATCCCATCCCGCCGAGACACCGGCGAG 120
Db 61 ATGCGATCCTGAGATACAGCGCCCTCATCCCATCCCGCCGAGACACCGGCGAG 120

Qy 121 CCCACTGTCGCCGCGACACATTAATCTGATCTCTACACAGCGCACTCGGAGCAGAG 180
Db 121 CCCACTGTCGCCGCGACACATTAATCTGATCTCTACACAGCGCACTCGGAGCAGAG 180

Qy 181 CGCTTATACAGCGCACAGCGCTCTCCGCGCACCGCCACACACAGATGATGCCGCCCG 240
Db 181 CGCTTATACAGCGCACAGCGCTCTCCGCGCACCGCCACACACAGATGATGCCGCCCG 240

Qy 241 ACCGAGCGCGCCGCGACACACACCTTCTGAAACACAGAAACAAAGTCCCGCCCAAG 300
Db 241 ACCGAGCGCGCCGCGACACACACCTTCTGAAACACAGAAACAAAGTCCCGCCCAAG 300

Qy 301 CGGCTGCATGTGTCACATCCCTTCCGCTTCCGGATCCAGACCTCCGACAAATGTTT 360
Db 301 CGGCTGCATGTGTCACATCCCTTCCGCTTCCGGATCCAGACCTCCGACAAATGTTT 360

Qy 361 GGCCAAATTTGGTAAATATTAGATGTTGAAATTTATTTTAATGAGCGGGCTCGAAGGGA 420
Db 361 GGCCAAATTTGGTAAATATTAGATGTTGAAATTTATTTTAATGAGCGGGCTCGAAGGGA 420

Qy 421 TTGGTTTCGTAACCTTCGAAATAGTCGGATCGGATCCGAGCTCCGACAAATGTCAC 480
Db 421 TTGGTTTCGTAACCTTCGAAATAGTCGGATCGGATCCGAGCTCCGAGCAATGTCAC 480

Qy 481 GGTACCGTGTAGAGCGCGTAAATTCAGGTTAATATCGACAGCAGCGTGATGACT 540
Db 481 GGTACCGTGTAGAGCGCGTAAATTCAGGTTAATATCGACAGCAGCGTGATGACT 540

Qy 541 AATAAAAGGCGGTGAACCCCTTACACCAATGCTCGAAATTAATCCAGTTGTGGCGCG 600
Db 541 AATAAAAGGCGGTGAACCCCTTACACCAATGCTCGAAATTAATCCAGTTGTGGCGCG 600

Qy 601 GTCTACAGCCCGACTTCTATGAGCGACGCTGCTGTTGTCAGCGCCACACGAGAGGA 660
Db 601 GTCTACAGCCCGACTTCTATGAGCGACGCTGCTGTTGTCAGCGCCACACGAGAGGA 660

Qy 661 TCTTCCATGTACAGTGGCCCGAGTTGCTGATATATCTTGTCAATGCTGCTTCCCA 720
Db 661 TCTTCCATGTACAGTGGCCCGAGTTGCTGATATATCTTGTCAATGCTGCTTCCCA 720

Qy 721 TATCCGGCGCGCACTGCTGACGTGCATACCGAGGGGCTCACCTTCGAGGGCGTGC 780
Db 721 TATCCGGCGCGCACTGCTGACGTGCATACCGAGGGGCTCACCTTCGAGGGCGTGC 780

Qy 781 ACCGTGTACACACCTTACAGCTGCGGCGCCCGACCCCGAATCCGCGCTATGCGCGA 840
Db 781 ACCGTGTACACACCTTACAGCTGCGGCGCCCGACCCCGAATCCGCGCTATGCGCGA 840

Qy 841 GTAGTGATCAGAGCCAGTGTATGCAATTAATGCTACAGGGGTTTACGCTGCATAC 900
Db 841 GTAGTGATCAGAGCCAGTGTATGCAATTAATGCTACAGGGGTTTACGCTGCATAC 900

Qy 901 CGCTAGCCCGAGCCACCCCTGCGACCTGCTGCTTACAGTGACAGTTAGCGAGGATT 960
Db 901 CGCTAGCCCGAGCCACCCCTGCGACCTGCTGCTTACAGTGACAGTTAGCGAGGATT 960

Qy 961 TATGCTGCCGACCCCTACACACACACTTGTCTCAGCCGCCACCTACGGGTTGGTGCC 1020
Db 961 TATGCTGCCGACCCCTACACACACACTTGTCTCAGCCGCCACCTACGGGTTGGTGCC 1020

Qy 1021 ATGAATGCTTTTGGCCCTTGGACCGATGCAAGACTAGGAGCCATGCTGATGTGGT 1080
Db 1021 ATGAATGCTTTTGGCCCTTGGACCGATGCAAGACTAGGAGCCATGCTGATGTGGT 1080

Qy 1081 CTGCTTCTTTTTCATTTGAGGCTAGTATATACCAAGGGGATACACCGTTTGTCTCCA 1140
Db 1081 CTGCTTCTTTTTCATTTGAGGCTAGTATATACCAAGGGGATACACCGTTTGTCTCCA 1140

Qy 1141 TATTAAATGATAAACCATTAACAAACAAAGCAAAACAAACAAACAAACAAACAA 1200
Db 1141 TATTAAATGATAAACCATTAACAAACAAAGCAAAACAAACAAACAAACAAACAA 1200

Qy 1201 ACCTTCAATGTGGGAGAGAGGAGCTTTCGAGGCCGAGTGTGCGACACATGCA 1260
Db 1201 ACCTTCAATGTGGGAGAGAGGAGCTTTCGAGGCCGAGTGTGCGACACATGCA 1260

Qy 1261 AGGACATCATTTCAGCACTCAAGAAACAAACAAACAAACAAACAAACAAACAA 1320
Db 1261 AGGACATCATTTCAGCACTCAAGAAACAAACAAACAAACAAACAAACAAACAA 1320

Qy 1321 GGCCGAGGGGTTCCGTAGA 1340
Db 1321 GGCCGAGGGGTTCCGTAGA 1340

RESULT 2
AF107204 1363 bp mRNA linear ROD 23-JUN-2000
LOCUS Mus musculus ataxin 2-binding protein (A2bp) mRNA, complete cds.
DEFINITION AF107204
ACCESSION AF107204
VERSION AF107204.1 GI:8671587
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Shibata.H., Huynh.D.P., Vo,T.T. and Pulst.S.-M.
TITLE A novel protein, A2BP, with RNA binding motif, binds to C-terminal ataxin-2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1363)
AUTHORS Shibata.H., Huynh.D.P., Vo,T.T. and Pulst.S.-M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1998) Division of Neurology and Neurogenetics Laboratory, Burns and Allen Research Institute, Cedars-Sinai Medical Center, UCLA School of Medicine, 8700 Beverly Blvd., Los Angeles, CA 90048, USA
FEATURES
source 1..1363
/organism="Mus musculus"
/db_xref="taxon:10090"
gene 1..1363
/gene="A2bp"
CDS 67..1203
/gene="A2bp"
/codon_start=1
/product="ataxin 2-binding protein"
/protein_id="AAF78292.1"
/db_xref="GI:8671588"

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 14:22:28 ; Search time 3582.04 Seconds
(without alignments)
10887.019 Million cell updates/sec

Title: US-09-809-545A-1
Perfect score: 1340
Sequence: 1 ggcgcgcgcctgacacaaat.....ggccgaagggttcgctaga 1340

Scoring table: OLIGO_NUC
Gapop_60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sta.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_in.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sta.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pin.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1340	100.0	1340	6	AX268800	AX268800 Sequence
2	194	14.5	1363	10	AF107204	AF107204 Mus muscu
3	194	14.5	2002	10	AB041596	AB041596 Mus muscu
4	168	12.5	115860	2	AC120661	AC120661 Rattus no
5	161	12.0	1586	10	AF191501	AF191501 Mus muscu
6	79	5.9	73929	2	AC103207	AC103207 Rattus no
7	70	5.2	1475	9	AB060859	AB060859 Macaca fa
8	70	5.2	1513	9	AK001027	AK001027 Homo sapi
9	70	5.2	2000	9	AF229057	AF229057 Homo sapi
10	70	5.2	2279	9	AF107203	AF107203 Homo sapi
11	70	5.2	2372	6	ARI34676	ARI34676 Sequence
12	70	5.2	2971	9	HSFOX14	AF109120 Homo sapi
13	70	5.2	3348	9	AF109106	AF109106 Homo sapi
14	70	5.2	164538	9	AC005774	AC005774 Homo sapi
15	67	5.0	115860	2	AC120661	AC120661 Rattus no
16	54	4.0	1547	9	AF094849	AF094849 Homo sapi
17	43	3.2	374	9	HSFOX08	AF109114 Homo sapi
18	43	3.2	1623	9	HSFOX08	AF109114 Homo sapi
19	40	3.0	234	9	HSFOX13	AF109119 Homo sapi
20	34	2.5	219	9	HSFOX05	AF109111 Homo sapi
21	34	2.5	112976	9	AC027683	AC027683 Homo sapi
22	34	2.5	169765	9	AC006075	AC006075 Homo sapi
23	32	2.4	629	9	HSFOX03	AF109109 Homo sapi
24	32	2.4	173824	2	AC095278	AC095278 Rattus no
25	31	2.3	169089	9	AC008537	AC008537 Homo sapi
26	31	2.3	169192	2	AC114672	AC114672 Mus muscu
27	31	2.3	217280	2	AC019337	AC019337 Homo sapi
28	31	2.3	224930	2	AC121911	AC121911 Mus muscu
29	31	2.3	235183	2	AC079427	AC079427 Mus muscu
30	30	2.2	5695	6	AX251292	AX251292 Sequence
31	30	2.2	5898	6	AX278016	AX278016 Sequence
32	30	2.2	5898	6	AX323715	AX323715 Sequence
33	30	2.2	5898	6	AX346842	AX346842 Sequence
34	30	2.2	6126	6	AX251328	AX251328 Sequence
35	30	2.2	6126	6	AX346733	AX346733 Sequence
36	30	2.2	6154	6	AX346207	AX346207 Sequence
37	30	2.2	6510	6	AX251401	AX251401 Sequence
38	30	2.2	6510	6	AX251810	AX251810 Sequence
39	30	2.2	6510	6	AX345296	AX345296 Sequence
40	30	2.2	57759	9	ALI37162	ALI37162 Human DNA
41	30	2.2	134787	2	AL356692	AL356692 Homo sapi
42	30	2.2	203279	9	AL354826	AL354826 Human DNA
43	30	2.2	260534	2	AC078912	AC078912 Mus muscu
44	29	2.2	219	9	HSFOX04	AF109110 Homo sapi
45	29	2.2	1076	9	BC020939	BC020939 Homo sapi

ALIGNMENTS

RESULT 1	AX268800	Sequence 1 from Patent WO0174901.	1340 bp	DNA	linear	PAT 29-OCT-2001
LOCUS	AX268800	Sequence 1 from Patent WO0174901.	1340 bp	DNA	linear	PAT 29-OCT-2001
DEFINITION	AX268800	Sequence 1 from Patent WO0174901.	1340 bp	DNA	linear	PAT 29-OCT-2001
ACCESSION	AX268800	Sequence 1 from Patent WO0174901.	1340 bp	DNA	linear	PAT 29-OCT-2001
VERSION	AX268800.1	GI:16541860				
KEYWORDS						
SOURCE						
ORGANISM						
		Norway rat.				
		Rattus norvegicus				
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;				
		Rattus.				
REFERENCE						
AUTHORS		Stanton,L.W. and White,R.T.				
TITLE		Secreted factors				

[illegible]

RESULT 14

```

US-09-789-561-55
;
; Sequence 55, Application US/09789561
; Patent No. US20020064818A1
;
; GENERAL INFORMATION:
;
; APPLICANT: Ni et al.
;
; TITLE OF INVENTION: 52 Human secreted proteins
;
; FILE REFERENCE: P2043P1
;
; CURRENT APPLICATION NUMBER: US/09/789,561
;
; CURRENT FILING DATE: 2001-02-22
;
; PRIOR APPLICATION NUMBER: PCT/US00/24008
;
; PRIOR FILING DATE: 2000-08-31
;
; PRIOR APPLICATION NUMBER: 60/152,317
;
; PRIOR FILING DATE: 1999-09-03
;
; PRIOR APPLICATION NUMBER: 60/152,315
;
; PRIOR FILING DATE: 1999-09-03
;
; NUMBER OF SEQ ID NOS: 194
;
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 55
;
; LENGTH: 1198
;
; TYPE: DNA
;
; ORGANISM: Homo sapiens
;
; US-09-789-561-55

```

	Query Match	3.7%	Score 49.2	DB 10	Length 1198
	Best Local Similarity 54.4%	Pred. No. 0.011			
	Matches 99	Conservative 0	Mismatches 83	Indels 0	Gaps 0
Qy	1137	TCCATATTAAATGATAAAACCATTTAAACAAACAAGCAAAAAACAAAAACAAAAA	1196		
Db	1011	TCAAAATAAATCTGGTAGCAACGTTTAAAAAATAAAAAAATAAAAAAATAAAAAA	1070		
Qy	1197	ACCAACCTTCCAAATGTGGGGAGAGAGGAGCTTTCCGAGGCCGCGAGTGTGGGACACATG	1256		
Db	1071	AAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA	1130		
Qy	1257	CAGTAGGACATCAGCTTTAGCACTCAAAGAAACAACGAAAAAATAAAAAAATAAAAAA	1316		
Db	1131	AAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA	1190		
Qy	1317	AA 1318			
Db	1191	AA 1192			

RESULT 15

```

RESUL 15
US-09-983-965-2109/c
; Sequence 2109, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26

```

```

, PRIOR APPLICATION NUMBER: US 09/465,231
, PRIOR FILING DATE: 1999-12-15
, PRIOR APPLICATION NUMBER: US 60/113,678
, PRIOR FILING DATE: 1998-12-17
, NUMBER OF SEQ ID NOS: 5912
, SEQ ID NO 2109
, LENGTH: 529
, TYPE: DNA
, ORGANISM: Bos taurus
, FEATURE:
, NAME/KEY: unseqre
, LOCATION: (335)
, OTHER INFORMATION:
, - OTHER INFORMATION: Clone ID: 34-LiB3057-015-Q1-K1-A6
US-09-983-965-2109

```

[illegible]

Search completed: March 15, 2003, 16:03:57
Job time : 132.943 secs

```
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4237 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-745-763-20

Query Match      3.9%; Score 52; DB 10; Length 4237;
Best Local Similarity 55.8%; Pred. No. 0.0038;
Matches 100; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 1150 ATAAACCATTAACAAACAAAGCAAAACAAACAAACAAACAAACAAACAAACCAACCTTCCAA 1209
DB 2430 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2489

QY 1210 TGTGGGAGAGAGAGAGCTTCCGAGCCGAGTGTTCGACACATGCATGATAGACATCA 1269
DB 2490 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2549

QY 1270 CTTTAGCACTCAAGAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACGCGAAGG 1329
DB 2550 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2609

RESULT 11
US-10-001-843-45
; Sequence 45, Application US/10001843
; Patent No. US20020132255A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Cafferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; CURRENT APPLICATION NUMBER: US/10/001,843
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,992
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45
; LENGTH: 1267
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (358)..(358)
; OTHER INFORMATION: a, c, g or t
; NAME/KEY: misc feature
; LOCATION: (478)..(478)
; OTHER INFORMATION: a, c, g or t
US-10-001-843-45

Query Match      3.7%; Score 49.8; DB 12; Length 1267;
Best Local Similarity 55.5%; Pred. No. 0.0077;
Matches 96; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 1146 AATGATAAACCATTAACAAACAAACAAACAAACAAACAAACAAACAAACAAACCAACCTT 1205
DB 976 AAAAAAAAAAAAAAAAAACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1035
```

```
QY 1206 CCAATGTGGGAGAGAGAGAGCTTTCGAGGCCGAGTGTTCGACACATGCATGATAGGAC 1265
DB 1036 CAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1095

QY 1266 ATCAGCTTTAGCACTCAAGAGAAACACGAAACAAACAAACAAACAAACAAACAAACAAAT 1318
DB 1096 AAAAAACAGAAACAAACAGATACAAACAAACAAACAAACAAACAAACAAACAAACAAAGAA 1148

RESULT 12
US-09-960-352-573/c
; Sequence 573, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 573
; LENGTH: 428
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 03-LIB3058-031-Q1-K1-A11
US-09-960-352-573

Query Match      3.7%; Score 49.4; DB 10; Length 428;
Best Local Similarity 54.7%; Pred. No. 0.0058;
Matches 98; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1140 ATATTAAATGATAAACCACTTTAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACC 1199
DB 407 ATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATATA 348

QY 1200 AACCTTCCATGTGGGAGAGAGAGCTTTCGAGCCGAGTGTTCGACACATGCAG 1259
DB 347 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAATATAAAAAAAAAAGAAATATAAAAAA 288

QY 1260 TAGGACATCACTTTAGCACTCAAGAGAAACAAACGAAACAAACAAACAAACAAACAAACAAAT 1318
DB 287 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATATAAAAAAAAAAAAAAAAAAAAAAAAAATAA 229

RESULT 13
US-09-960-352-3400/c
; Sequence 3400, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 3400
; LENGTH: 446
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 15-LIB3058-052-Q1-K1-D11
US-09-960-352-3400

Query Match      3.7%; Score 49.4; DB 10; Length 446;
Best Local Similarity 54.7%; Pred. No. 0.006;
Matches 98; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
```


Db 289 TNGTNNNTNTTATNTTTCTTTCTTTCTTTTGAATTTAAANAAAGGNAAAAA 230

Qy 1131 TTTTGTCCATATTAAATGATAAACCAATTAACAAACAGCAAAAAACAAAAAC 1190

Db 229 ANAANTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 170

Qy 1191 AAAAAAACCACTTCCAATGTGGGAGAGAGAGCTTTCCGAGGCCGAGTTGCGA 1250

Db 169 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 110

Qy 1251 CACATGCGAGTAGGCATCATTAGCACTCAAGAGACACAGAAAAAATAAAAAA 1310

Db 109 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 50

Qy 1311 AAAAAATAA 1318

Db 49 AAAAAAAA 42

RESULT 8

US-09-825-294-53/c

; Sequence 53 Application US/09825294

; Patent No. US20020004491A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Stolk, John A.

; APPLICANT: Algate, Paul A.

; APPLICANT: Fling, Steven P.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE

; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER

; FILE REFERENCE: 210121.484CS

; CURRENT APPLICATION NUMBER: US/09/825.294

; CURRENT FILING DATE: 2001-04-03

; NUMBER OF SEQ ID NOS: 215

; SOFTWARE: FastSEQ for Windows Version 3.0

; SEQ ID NO 53

; LENGTH: 396

; TYPE: DNA

; ORGANISM: Homo sapien

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)...(396)

; OTHER INFORMATION: n = A,T,C or G

US-09-825-294-53

Query Match 3.9%; Score 52.8; DB 10; Length 396;

Best Local Similarity 48.4%; Pred. No. 0.00079;

Matches 120; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

Qy 1071 TCATGTGGGTCTCGTCTCTTTCTTTCTTTCAGGCTAGTATATACCAAGGGGGATACACCG 1130

Db 289 TNGTNNNTNTTATNTTTCTTTCTTTTGAATTTAAANAAAGGNAAAAA 230

Qy 1131 TTTTGTCCATATTAAATGATAAACCAATTAACAAACAGCAAAAAACAAAAAC 1190

Db 229 ANAANTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 170

Qy 1191 AAAAAAACCACTTCCAATGTGGGAGAGAGAGCTTTCCGAGGCCGAGTTGCGA 1250

Db 169 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 110

Qy 1251 CACATGCGAGTAGGCATCATTAGCACTCAAGAGAAACAAAGAAAAAATAAAAAA 1310

Db 109 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 50

Qy 1311 AAAAAATAA 1318

Db 49 AAAAAAAA 42

RESULT 9

US-09-960-352-4582/c

; Sequence 4582, Application US/09960352

; Patent No. US20020137139A1

; GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.

; APPLICANT: Tao, Nengbing

; APPLICANT: Byatt, John C.

; APPLICANT: Mathialagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION

; FILE REFERENCE: 16511.006/37-21(10298)C

; CURRENT APPLICATION NUMBER: US/09/960,352

; CURRENT FILING DATE: 2001-09-24

; NUMBER OF SEQ ID NOS: 15112

; SEQ ID NO 4582

; LENGTH: 393

; TYPE: DNA

; ORGANISM: Bos taurus

; OTHER INFORMATION: Clone ID: 20-LIB3057-013-Q1-K1-E7

US-09-960-352-4582

Query Match 3.9%; Score 52; DB 10; Length 393;

Best Local Similarity 53.4%; Pred. No. 0.0013;

Matches 109; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

Qy 1122 ATACAACCGTTTGTCTCATATTAAATGATAAAACCAATTAACAAACAGCAAAAAACAA 1181

Db 213 ATAAAAAATTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 154

Qy 1182 AACAAAAACAAAAAACCAACCTTCCAATGTGGGAGAGAGAGCTTTCCGAGGCCGA 1241

Db 153 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 94

Qy 1242 GTGTGGCAGCATGCGAGTAGGCATCATTAGCACTCAAGAGAAACAAAGAAAAA 1301

Db 93 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 34

Qy 1302 AAAAAAATAAAAAAATAAAGCGCGC 1325

Db 33 AAAAAAATAAAAAAATAAAGGCGG 10

RESULT 10

US-09-745-763-20

; Sequence 20, Application US/09745763

; Patent No. US20020065394A1

; GENERAL INFORMATION:

; APPLICANT: Jacobs, Kenneth

; McCoy, John M.

; LaVallie, Edward R.

; Collins-Racie, Lisa A.

; Evans, Cheryl

; Marberg, David

; Treacy, Maurice

; Spaulding, Vikki

; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

; NUMBER OF SEQUENCES: 219

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetics Institute, Inc.

; STREET: 87 CambridgePark Drive

; CITY: Cambridge

; STATE: MA

; COUNTRY: U.S.A.

; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/745,763

; FILING DATE: 18-Jun-2000

; CLASSIFICATION: <Unknown>

; ATTORNEY/AGENT INFORMATION:

Qy	5	CCGCCCTGTGACACAATGGCTTAG-CTTATGGCTCAGCGCAGTTGCTCTCACCCCGAAGTG	63
Db	1	CTGCCCTGTGACACAATGGCTCAGCTTACGCTTCGGCCCAAGTTGCTCCCGCAGAACG	60
Qy	64	GCATCTCT-GCAGAATACACGGCCCTCATCCCCATCCGGCCGACAG-ACACGGCCGACG	121
Db	61	GTATCCCGCGGGAATACACGGCCCTCATCCCCACCCGCGCAGAGTACACAGGCCAG	120
Qy	122	CCACTGTCCCCG-CCACACATTTAAACTTG-ATCCTCTTACACAGACGCACTCGGACGAGA	179
Db	121	CCACGGTTCGGACGACATTTAACTGTATACCTCCGCGCCAGACGCACTCCGACGAGA	180
Qy	180	GC---GTTTATACAAGGCGACAG-CCGCTCTCGGSCACCGCCACA	219
Db	181	GCCCGGCGGACACAGCGCTCAGACCGCTCTTGACACCGCCACA	224

RESULT 6
US-09-864-761-5893
; Sequence 5893, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864, 761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 5893
; LENGTH: 454
; TYPE: DNA
; ORGANISM: Homo sapiens

```

; FEATURE:
; OTHER INFORMATION: MAP TO AC006075.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
US-09-864-761-5893

Query Match          6.4%; Score 86.4; DB 10; Length 454;
Best Local Similarity 83.5%; Pred. No. 3.3e-12;
Matches 157; Conservative 0; Mismatches 26; Indels 5; Gaps

Qy 1 GCGGCGCGCCCTGACACATGGCTCAG-CTTATGCTCTCAGCGCAGTTCGCTCCACCCCGAG 59
Db 246 GCCGCTGCCCTGACACATGGCTCAGCCTTAGCCTTCGCGCCAGTTTGCTCCCGCCGCGAG 305

Qy 60 AATGGCATCTCT-CGAAGATACAGGCGCCCTCATCCCATCCGCGCCAGAG-ACACGGCG 117
Db 306 AACGGTATCCCGCGGAATACACGGCGCCCTCATCCCGCCGCGCGCAGAGTACACAGCG 365

Qy 118 CAGCGCCACTGTCCCGG-CCACACATTAAACTTG-ATCTCTCTTACACAGACGCACTCGGAG 175
Db 366 CAGACACAGGTTCCGANGACACATTAAACCTGTACCTTCCGCCCCAGACGACTCCGAG 425

Qy 176 CAGAGCGC 183
Db 426 CAGAGGCC 433

```

```

RESULT 7
US-09-970-966-53/c
; Sequence 53, Application US/09970966
; Patent No. US20020173638A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Molesh, David Alan
; APPLICANT: Fling, Steven P.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C6
; CURRENT APPLICATION NUMBER: US/09/970,966
; CURRENT FILING DATE: 2001-10-02
; NUMBER OF SEQ ID NOS: 215
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 53
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 234, 225, 228, 235, 240, 246, 257, 266, 274, 279, 281, 282,
; LOCATION: 283, 285, 287, 288, 290, 291, 292, 293, 294, 295, 296, 297,
; LOCATION: 300, 303, 307, 311, 313, 314, 317, 318, 319, 320, 321,
; LOCATION: 323, 324, 328, 329, 330, 336, 337, 338, 339, 340, 341
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc_feature
; LOCATION: 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 356,
; LOCATION: 357, 358, 359, 362, 363, 364, 365, 366, 367, 373, 380, 381,
; LOCATION: 382, 385, 387, 388, 389, 390, 392
; OTHER INFORMATION: n = A,T,C or G
US-09-970-966-53

```

Qy 1071 TGATGTGGGTCTCGTTCTTTCTTCATTGCAGGCTAGTATATACCAAGGGGGATACAACCG 1130

[illegible]

Query Match	8.2%	Score 109.6;	DB 10;	Length 125;
Best Local Similarity	92.7%	Pred. No. 2.8e-18;		
Matches 115;	Conservative	0;	Mismatches 9;	Indels 0;
Gaps				

QY 239 CGACCGGCGCCAGCCCCAGACACAACTTCTGAAACACAGAAACAAGTCCCAGCCCA 298

Db
1 CGACGGATGGCCAGCCCCAGACACAACTTCTGAAAAACACGGAAAAACAAGTCTCAGCCCA 60

QV 299 AGCGGCTGCATGTGTCCAACATCCCCCTTCCGGTTCGGGATCCAGACCTCCGACAAATGT 358

Db 61 AGCGGCTGCATGTCCTCCAATATCCCCCTTCAGGTTCCGGGATCCGGACCTCAGACAAATGT 120

Ov 359 TTGG 362

D_b 121 TGG 124

RESULT 5

US-09-864-761-22653

```

; Sequence 22653, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzei, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/006666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 22653
; LENGTH: 224
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006075.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
; OTHER INFORMATION: NT HIT: g18922072, EVALU6 1.00e-123
; OTHER INFORMATION: SWISSPROT HIT: Q09457, EVALU6 8.00e-03
; OTHER INFORMATION: EST_HUMAN HIT: AU118435.1, EVALU6 1.00e-123
; US-09-864-761-22653

```

Query Match 6.6%; Score 88; DB 10; Length 224;
Best Local Similarity 82.6%; Pred. No. 9.4e-13;
Matches 185; Conservative 0; Mismatches 30; Indels 9; Gaps 7;

Qy 301 CGGCTGCATGTGTCAACATCCCTTCCGTTCCGGATCCAGACCTCCGACAAATGTTT 360
Db 301 CGGCTGCATGTGTCAACATCCCTTCCGTTCCGGATCCAGACCTCCGACAAATGTTT 360
Qy 361 GGCCAATTTGTAATAATTAGATGTTGAAATTTATTTAATAGCGGGCTCGAAGGA 420
Db 361 GGCCAATTTGTAATAATTAGATGTTGAAATTTATTTAATAGCGGGCTCGAAGGA 420
Qy 421 TTTGGTTTTCGTAACCTTCGAAATAGTCGGATCGGACAGCGGCGAGGAGAAATTGCAC 480
Db 421 TTTGGTTTTCGTAACCTTCGAAATAGTCGGATCGGACAGCGGCGAGGAGAAATTGCAC 480
Qy 481 GGTACCGTGTAGAGGGCCGTAATAATCGAGGTTAATAATGCGACAGCAGCGGTGATGACT 540
Db 481 GGTACCGTGTAGAGGGCCGTAATAATCGAGGTTAATAATGCGACAGCAGCGGTGATGACT 540
Qy 541 AATAAAGAGCGGTGAACCCCTACACCAATGCTGGAATTAATTCAGTTGTGGGCGG 600
Db 541 AATAAAGAGCGGTGAACCCCTACACCAATGCTGGAATTAATTCAGTTGTGGGCGG 600
Qy 601 GTCTACAGCCCGACTTCTATGACGACCGTGTGTTGTGCGAGGCCAACAGGAGGA 660
Db 601 GTCTACAGCCCGACTTCTATGACGACCGTGTGTTGTGCGAGGCCAACAGGAGGA 660
Qy 661 TCTTCATGTACAGTGGCCCGACTTCTATGATATATCTTCTGCAATGCTGGCTTCCA 720
Db 661 TCTTCATGTACAGTGGCCCGACTTCTATGATATATCTTCTGCAATGCTGGCTTCCA 720
Qy 721 TATCGGGCGGACGTGTGAGTGCATACCGAGGGGCTCACCTTCGAGGGCGTGGTGC 780
Db 721 TATCGGGCGGACGTGTGAGTGCATACCGAGGGGCTCACCTTCGAGGGCGTGGTGC 780
Qy 781 ACCGTGTACACACCTTCAGAGCTCGCGGCCGCCCAATCCCGGCTATGGCGGA 840
Db 781 ACCGTGTACACACCTTCAGAGCTCGCGGCCGCCCAATCCCGGCTATGGCGGA 840
Qy 841 GTAGTGTATCAAGACCGAGTGTATGGCAATTAATTTGCTACAGGGTGTGCTGCATAC 900
Db 841 GTAGTGTATCAAGACCGAGTGTATGGCAATTAATTTGCTACAGGGTGTGCTGCATAC 900
Qy 901 CGCTACGGCCAGCCACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Db 901 CGCTACGGCCAGCCACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Qy 961 TATGCTGCGACCCCTACACACACACTTGTCTCAGGCCCCACCTACGGGCTTGGTGC 1020
Db 961 TATGCTGCGACCCCTACACACACACTTGTCTCAGGCCCCACCTACGGGCTTGGTGC 1020
Qy 1021 ATGAATGCTTTTGGCCCTTGAACCGATGCAAGACTAGGAGCCATGCTGATGATGGGT 1080
Db 1021 ATGAATGCTTTTGGCCCTTGAACCGATGCAAGACTAGGAGCCATGCTGATGATGGGT 1080
Qy 1081 CTCGTTCTTTCTTCAATGAGGCTAGTATATACCAAGGGGGATACAAACGTTTCTCTCA 1140
Db 1081 CTCGTTCTTTCTTCAATGAGGCTAGTATATACCAAGGGGGATACAAACGTTTCTCTCA 1140
Qy 1141 TATTAAATGATAAACCATTAAACAAACAGCAAAACAAACAAACAAACAAACAAACAA 1200
Db 1141 TATTAAATGATAAACCATTAAACAAACAGCAAAACAAACAAACAAACAAACAAACAA 1200
Qy 1201 ACCTTCCAATGTGGGAGAGAGAGCTTTCGAGGGCCCGAGTGTGCGACACATGCACT 1260
Db 1201 ACCTTCCAATGTGGGAGAGAGAGCTTTCGAGGGCCCGAGTGTGCGACACATGCACT 1260
Qy 1261 AGGACATCATTTAGCACTCAAGAAACAAACGAAAAAATAAGC 1320
Db 1261 AGGACATCATTTAGCACTCAAGAAACAAACGAAAAAATAAGC 1320
Qy 1321 GGCAGAGGGTTTCCTAGA 1340
Db 1321 GGCAGAGGGTTTCCTAGA 1340

RESULT 2
US-09-794-591-1
; Sequence 1, Application US/09794591
; Patent No. US20010018198A1
; GENERAL INFORMATION:
; APPLICANT: Pulst, Stefan M.
; TITLE OF INVENTION: Nucleic Acids Encoding Ataxin-2 Binding Proteins,
; TITLE OF INVENTION: Products Related Thereto and Methods of Using Same
; FILE REFERENCE: CR 3093
; CURRENT APPLICATION NUMBER: US/09/794,591
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/145,391
; PRIOR FILING DATE: 1998-09-01
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2372
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (987) .. (1979)
US-09-794-591-1
Query Match 64.0%; Score 857.2; DB 10; Length 2372;
Best Local Similarity 82.7%; Pred. No. 7.5e-205;
Matches 1141; Conservative 0; Mismatches 138; Indels 101; Gaps 10;
Qy 1 GCGGCGGCGCTGACAAATGGCTCAG-CTTATGCTTCAGGCGAGTTGCTTCACCCCGAG 59
Db 1029 GCGGCTGCGCTGACAAATGGCTCAGCTTTACGCTTCGGCCAGTTTGTCTCCCGCGAG 1088
Qy 60 AATGGCATCT-GCAGAATACACGGCCCTCATCCCATCCGCGCCAGAG-ACACCGGC 117
Db 1089 AACGGTATCCCGCGGAATACACGGCCCTCATCCCATCCGCGCCAGAGTACACAGC 1148
Qy 118 CAGCCCACTGTCCCG-CCACACATTAACCTTG-ATCCTCTACACAGACGACCTCGGAG 175
Db 1149 CAGACACGGTTCGAGGACACATTAACCTGTACCTCCCGCCAGACGACCTCCGAG 1208
Qy 176 CAGAGC--GCTTATACAAAGGCGACAG-CGGTCTCCGCGACCGGCGACACAGACGATGAT 231
Db 1209 CAGAGCCGCGCGACAGAGCGCTCAGACCGCTCTCTGSCACCGGCGACACAGACGATGAC 1268
Qy 232 GCGCCCGCGACGCGCGCCAGCCAGACACAACTTCTGAAACACAGAAACAAAGTCC 291
Db 1269 GCAGCACGCGGATGGCGAGCCCGCCAGACACAACTTCTGAAACACAGAAACAAAGTCT 1328
Qy 292 CAGCCCAAGCGCTGCATGTGTCCAAACATCCCTTCCGGTTCGGGATCCAGACCTCCGA 351
Db 1329 CAGCCCAAGCGCTGCATGTCTCCAAATATCCCTTCAGTTCCGGATCCGACCTCAGA 1388
Qy 352 CAAATGTTTGGCCAAATTTGGTAAATATATAGATGTTGAAATATTTTAAATAGCGGGGC 411
Db 1389 CAAATGTTTGGTCAATTTGGTAAATATCTAGATGTTGAAATATTTTAAATAGCGGGGC 1448
Qy 412 TCGAGGGATTTGCTTTCGTAACCTTCGAAATAGTCGGATCGGACAGCGCGAGGAG 471
Db 1449 TCAAGGGATTTGCTTTCGTAACCTTCGAAATAGTCGGATCGGACAGCGCGAGGAG 1508
Qy 472 AAATTGACCGGTACCGTGTAGAGGGCGGTAAATTCGAGGTAAATTAATTCGACAGCAGC 531
Db 1509 AAATTACGCGTACCGTGTAGAGGGCGGTAAATTCGAGGTAAATTAATTCGACAGCAGC 1568
Qy 532 GTGATGACTAATAAAAAAGCGGTGAACCCCTACACCAATGCTGGAATTAATTCAGTT 591
Db 1569 GTAATGACAAATAAAAAAGCGGTCAACCCCTTATACCAATGCTGGAATTAATTCAGTT 1628
Qy 592 GTGGCGGGTCTACAGCCCGGACTTCTATGAGGACGCGTGTGTTGCGGAGGCCAAC 651
Db 1629 GTGGGTGAGTCTACAGTCCCGAATTTCTATGAGGACGCGTCTGTTGTGCGAGGCCAAC 1688

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 12:44:58 ; Search time 118.943 Seconds
(without alignments)
7907.139 Million cell updates/sec

Title: US-09-809-545A-1

Perfect score: 1340

Sequence: 1 gggcgccgcctgacacaat.....ggccgaagggttcgctaga 1340

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 501302 seqs, 35092545 residues

Total number of hits satisfying chosen parameters: 1002604

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1340	100.0	1340	10	US-09-809-545A-1
2	857.2	64.0	2372	10	US-09-794-591-1
3	123.8	9.2	473	10	US-09-864-761-11804
4	109.6	8.2	125	10	US-09-864-761-28375
5	88	6.6	224	10	US-09-864-761-28653
6	86.4	6.4	454	10	US-09-864-761-5893
C 7	52.8	3.9	396	9	US-09-970-966-53
C 8	52.8	3.9	396	10	US-09-825-294-53
C 9	52	3.9	4237	10	US-09-745-763-20
10	52	3.9	1267	12	US-10-001-843-45
11	49.8	3.7	428	10	US-09-960-352-573
C 12	49.4	3.7	446	10	US-09-960-352-3400
C 13	49.4	3.7	1198	10	US-09-789-561-55
14	49.2	3.7	529	10	US-09-983-965-2109
C 15	48.8	3.6	415	10	US-09-960-352-2223
C 16	48.4	3.6	200	10	US-09-960-352-4816
C 17	47.8	3.6	312	10	US-09-960-352-8414
C 18	47.8	3.6	425	10	US-09-960-352-4010
C 19	47.8	3.6			

C 20	47.6	3.6	2305	10	US-09-795-006A-124	Sequence 124, App
C 21	47.4	3.5	217	10	US-09-960-352-6539	Sequence 6539, App
C 22	47.4	3.5	283	10	US-09-960-352-9095	Sequence 9095, App
C 23	47.2	3.5	259	10	US-09-960-352-7233	Sequence 7233, App
C 24	47.2	3.5	463	10	US-09-960-352-7186	Sequence 7186, App
C 25	47	3.5	237	10	US-09-960-352-14543	Sequence 14543, A
C 26	46.8	3.5	241	10	US-09-960-352-12302	Sequence 12302, A
C 27	46.6	3.5	242	10	US-09-960-352-3873	Sequence 3873, App
C 28	46.2	3.4	207	10	US-09-960-352-3946	Sequence 3946, App
C 29	46.2	3.4	344	10	US-09-960-352-1036	Sequence 1036, App
C 30	46.2	3.4	439	10	US-09-960-352-1009	Sequence 1009, App
C 31	46.2	3.4	491	10	US-09-878-574-4301	Sequence 4301, App
C 32	46.2	3.4	516	10	US-09-960-352-5785	Sequence 5785, App
C 33	46.2	3.4	4167	10	US-09-764-878-282	Sequence 282, App
C 34	46.2	3.4	4167	10	US-09-764-860-1145	Sequence 1145, App
C 35	46.2	3.4	4167	10	US-09-764-846-345	Sequence 345, App
C 36	46	3.4	202	10	US-09-764-846-280	Sequence 280, App
C 37	46	3.4	936	10	US-09-739-907-47	Sequence 47, Appl
C 38	45.6	3.4	390	10	US-09-960-352-1976	Sequence 1976, App
C 39	45.6	3.4	831	10	US-09-729-674-67	Sequence 67, Appl
C 40	45.4	3.4	430	10	US-09-960-352-9747	Sequence 9747, App
C 41	45.2	3.4	545	10	US-09-878-574-4299	Sequence 4299, App
C 42	45.2	3.4	741	10	US-09-910-943-432	Sequence 432, App
C 43	45.2	3.4	758	10	US-09-910-943-490	Sequence 490, App
C 44	45.2	3.4	760	10	US-09-910-943-366	Sequence 366, App
C 45	45	3.4	448	10	US-09-878-574-4308	Sequence 4308, App

ALIGNMENTS

RESULT 1

US-09-809-545A-1
; Sequence 1, Application US/09809545A
; Patent No. US20020110804A1
; GENERAL INFORMATION:
; APPLICANT: Stanton, Lawrence W.
; APPLICANT: White, R. Tyler
; TITLE OF INVENTION: SECRETED FACTORS
; FILE REFERENCE: SCIOS.017A
; CURRENT APPLICATION NUMBER: US/09/809,545A
; CURRENT FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1340
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-09-809-545A-1

Query Match 100.0%; Score 1340; DB 10; Length 1340;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	CGCGCGCCCTTGACACAAATGGCTCAGCTTATCGCTCAGCGCAGTTGCTTCCACCCAG	60
Db	1	CGCGCGCCCTTGACACAAATGGCTCAGCTTATCGCTCAGCGCAGTTGCTTCCACCCAG	60
Qy	61	ATCGCATCTTCGAGATACACGGCCCTCATCCCATCCGCGCCAGACACCGGCCAG	120
Db	61	ATCGCATCTTCGAGATACACGGCCCTCATCCCATCCGCGCCAGACACCGGCCAG	120
Qy	121	CCCACTGTCTCCCGCCACACATTAACTTGTATCTCTTACACAGACGCACTCGGAGCAG	180
Db	121	CCCACTGTCTCCCGCCACACATTAACTTGTATCTCTTACACAGACGCACTCGGAGCAG	180
Qy	181	CGCTTATACAGCGCAGCGCTCTCGGCAACCGCCACACAGACAGATGATCGGCCCG	240
Db	181	CGCTTATACAGCGCAGCGCTCTCGGCAACCGCCACACAGACAGATGATGCGGCCCG	240
Qy	241	ACGAGCGCCAGCCCGCCACACAACTTCTGAAAAACACAGAAAAACAGTCCCGCCCAAG	300
Db	241	ACGAGCGCCAGCCCGCCACACAACTTCTGAAAAACACAGAAAAACAGTCCCGCCCAAG	300

```
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:DNA
US-09-331-581-3

Query Match      3.2%; Score 42.4; DB 3; Length 7286;
Best Local Similarity 58.9%; Pred. No. 0.068;
Matches 73; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 1139 CATATTAAATGATAAACCATTAACAAACAAAGCAAGCAAAACAAACAAACAAACAAACAAAC 1198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4613 CAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4672

Qy 1199 CAACCTTCCAATGTGGGAGAGAGAGCACTTCCGAGGCCCGAGTGTTCGACACATGCA 1258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4673 AAAAAGTACCTTCTGGAGCGGAAGAACCCAGCCGGATCCAGACATGATAAGATACATTGA 4732

Qy 1259 GTAG 1262
    ||
Db 4733 TGAG 4736

RESULT 15
US-09-331-581-14
; Sequence 14, Application US/09331581
; Patent No. 6130070
; GENERAL INFORMATION:
; APPLICANT: TOHDA, Hideki
; APPLICANT: HAMA, Yuko
; APPLICANT: KUMAGAI, Hiromichi
; TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
; TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
; TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF
; FILE REFERENCE: 0059-1142-0PCT
; CURRENT APPLICATION NUMBER: US/09/331,581
; CURRENT FILING DATE: 1999-06-30
; EARLIER APPLICATION NUMBER: PCT/JP98/04929
; EARLIER FILING DATE: 1998-10-30
; EARLIER APPLICATION NUMBER: JP 9-314608
; EARLIER FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 14
; LENGTH: 7938
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:DNA
US-09-331-581-14

Query Match      3.2%; Score 42.4; DB 3; Length 7938;
Best Local Similarity 58.9%; Pred. No. 0.072;
Matches 73; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 1139 CATATTAAATGATAAACCATTAACAAACAAAGCAAGCAAAACAAACAAACAAACAAAC 1198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5265 CAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 5324

Qy 1199 CAACCTTCCAATGTGGGAGAGAGAGCACTTCCGAGGCCCGAGTGTTCGACACATGCA 1258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5325 AAAAAGTACCTTCTGGAGCGGAAGAACCCAGCCGGATCCAGACATGATAAGATACATTGA 5384

Qy 1259 GTAG 1262
    ||
Db 5385 TGAG 5388
```

Search completed: March 15, 2003, 16:01:35
Job time : 134.316 secs

Query Match 3.4%; Score 45.2; DB 4; Length 2394;
Best Local Similarity 47.8%; Pred. No. 0.0058;
Matches 131; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 3 GGCGCCCTGACACATGGCTGCTTATCGCTCAGCGCAGTTGCTCCACCCAGAAAT 62
DB 1742 GGTGCTCCCATACCAACCTTACCAACCCCGGAGCTGATGATGTGAGCTGTG 1801

QY 63 GGCAATCTGTGCAAGATACACGGCCCTCATCCCATCCCGCCAGAGACACCGGCCAGCC 122
DB 1802 GCGGCTGTGCTGCTGCTCCCACTTACCAAGCGCTCCCGCCAGCACCGGCTGCT 1861

QY 123 CACTGTCCCGCCGACACATTAATCTGATCTCTTACACAGCAGCTCGGAGCAGAGCG 182
DB 1862 CAGCCCTCGGACTCGGATGACTGGAGTCTGTCACCCCTCCCGCTCTGATGACCCCTG 1921

QY 183 CTTATACAGCGCAGCGCTCGGACCGCCACACAGACAGATGATGCGGCCGCGAC 242
DB 1922 CCACTCTGGGCGCTTACCTGGCTCCGTTGTTGTTATCCCGAGCAGCAGCAGCTGTG 1981

QY 243 CGAGCGCCAGCCCGACACACAACTTCTGAAAC 276
DB 1982 ACCCTGCCCCCTCCCTGGGCCCCCTCTGAAAGC 2015

RESULT 10
US-08-676-967-2
; Sequence 2, Application US/08676967
; Patent No. 5747317
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,967
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-676-967-2

Query Match 3.4%; Score 45; DB 1; Length 2277;
Best Local Similarity 32.4%; Pred. No. 0.0064;
Matches 109; Conservative 46; Mismatches 178; Indels 3; Gaps 1;

QY 215 CCACACAGACAGATGCTCCGCCCCGAGCGGCCGAGCCAGCCAGACACCTTCTGAAA 274
DB 257 CNAARGAARAGGNAARAAAYGARAAYWSNGARTGYCCNNAARAAARGARCCNNAARAA 316

QY 275 ACACAGAAACAAAGTCCAGCGCCGCTGATGTGTCCAAACATCCCTTCCGGTTCC 334
DB 317 ARGCUAARGTNGCNGAYAAARAGCGNMGNYTNATHATGMAAAYTNTNSNTTAAATGYW 376

QY 275 ACACAGAAACAAAGTCCAGCGCCGCTGATGTGTCCAAACATCCCTTCCGGTTCC 334
DB 317 ARGCUAARGTNGCNGAYAAARAGCGNMGNYTNATHATGMAAAYTNTNSNTTAAATGYW 376

QY 335 GGGATCCAGACCTCCGACAAATGTTTGGCCAAATTTGGTAAATAATTAGATGTTGAAATTA 394
DB 377 SNGARGAYAYTNAARACNGTNTTYGCNCARTTYGGNGCNGTNYNGARGTNAAYATHC 436

QY 395 TTTTAA---TGAGCGGGGCTCGAAGGATTTGGTTTCGTAACCTTCGAAAATAGTCGGG 451
DB 437 CNMGNAACNGAYGGAARATGNGNGNTTYGNTTYGTNCARTTYAARAAAYTNTYNG 496

QY 452 ATCGGACAGCGGAGGAGAAATTCACGGTACCGTGGTAGAGGGCGGTAARATCCGAGG 511
DB 497 ARGCGNNAARGCNYTNAARGGNATGAAYATGAARGARATHAARGGNMGNACNGTNGCG 556

QY 512 TTAATAATCGACAGCAGCGCTGATGACTAATAAAA 547
DB 557 TNGAYTGGGCTGNGCNAARGAYAAARTAYAAARGAYA 592

RESULT 11
US-08-676-974-2
; Sequence 2, Application US/08676974
; Patent No. 5770422
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,974
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-676-974-2

Query Match 3.4%; Score 45; DB 1; Length 2277;
Best Local Similarity 32.4%; Pred. No. 0.0064;
Matches 109; Conservative 46; Mismatches 178; Indels 3; Gaps 1;

QY 215 CCACACAGACAGATGCTCCGCCCCGAGCGGCCGAGCCAGCCAGACACCTTCTGAAA 274
DB 257 CNAARGAARAGGNAARAAAYGARAAYWSNGARTGYCCNNAARAAARGARCCNNAARAA 316

QY 275 ACACAGAAACAAAGTCCAGCGCCGCTGATGTGTCCAAACATCCCTTCCGGTTCC 334
DB 317 ARGCUAARGTNGCNGAYAAARAGCGNMGNYTNATHATGMAAAYTNTNSNTTAAATGYW 376

Db 926 ACCCTGCCCTCCCTGGGCCCCCTCTGAAAGC 959

RESULT 7

US-09-271-815-3
; Sequence 3, Application US/09271815
; Patent No. 6297036
; GENERAL INFORMATION:
; APPLICANT: BERGSMAN, DERK
; APPLICANT: SHABON, USMAN
; TITLE OF INVENTION: A YAK-1 RELATED SERINE/THREONINE PROTEIN KINASE-HTLAR33
; FILE REFERENCE: GH-70172-1
; CURRENT APPLICATION NUMBER: US/09/271,815
; CURRENT FILING DATE: 1999-03-18
; EARLIER APPLICATION NUMBER: 09/027,064
; EARLIER FILING DATE: 1998-02-20
; EARLIER APPLICATION NUMBER: 60/053,924
; EARLIER FILING DATE: 1997-07-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1338
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-271-815-3

Query Match 3.4%; Score 45.2; DB 4; Length 1338;
Best Local Similarity 47.8%; Pred. No. 0.0041;
Matches 131; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 3 GGCGCGCCCTGACACAAATGGCTCAGCTTATGCTCAGCGAGTTGCTCCACCCCAAGAAAT 62
Db 686 GTCTGCTCCCATCACCAACCTCACCAACCCCGAGCTGATGATGTGAGCCTGGTG 745

QY 63 GGCATCTCCGAGAAATACACGGCCCTCATCCCATCCCGCGCCAGAGACACCGGCCAGCC 122
Db 746 GGGCGGCTGTGACTGTCTCCCACTCACCAAGCGCTGCGCCCGCCAGCACCGCGCTGCT 805

QY 123 CACTGTCCCGCCACACATTAACCTTATGCTCTACACAGCGCACTCGGAGCAGAGCG 182
Db 806 CAGCCCTCCGGACTCGGATGACTGGAGTGTCTCCACCCCTCCCGCTCTCTGATGACCC 865

QY 183 CTATATAAGCGACAGCCGCTCTCCGCGCCGCGCACACAGACAGATGATGCGCGCCCGAC 242
Db 866 CCACTCTGGGCGCTCCTGCGCTCTGCTGTGTACCCACAGACAGACAGCAGCAGCTGCTG 925

QY 243 CGACGGCCAGCCCGACACAACTTCTGAAAC 276
Db 926 ACCCTGCCCTCCCTGGGCCCCCTCTGAAAGC 959

RESULT 8

US-09-027-064-1
; Sequence 1, Application US/09027064
; Patent No. 6133006
; GENERAL INFORMATION:
; APPLICANT: SHABON, USMAN
; APPLICANT: BERGSMAN, DERK
; TITLE OF INVENTION: A YAK-1 RELATED SERINE/THREONINE
; TITLE OF INVENTION: PROTEIN KINASE-HTLAR33
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/027,064
; FILING DATE: 20-FEB-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/053,924
; FILING DATE: 28-JUL-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH-70172
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2394 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-027-064-1

Query Match 3.4%; Score 45.2; DB 3; Length 2394;
Best Local Similarity 47.8%; Pred. No. 0.0058;
Matches 131; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 3 GGCGCGCCCTGACACAAATGGCTCAGCTTATGCTCAGCGAGTTGCTCCACCCCAAGAAAT 62
Db 1742 GGTCTGCTCCCATCACCAACCTCACCAACCCCGAGCTGATGATGTGAGCCTGGTG 1801

QY 63 GGCATCTCCGAGAAATACACGGCCCTCATCCCATCCCGCGCCAGAGACACCGGCCAGCC 122
Db 1802 GGGCGGCTGTGACTGTCTCCCACTCACCAAGCGCTGCGCCCGCCAGCACCGCGCTGCT 1861

QY 123 CACTGTCCCGCCACACATTAACCTTATGCTCTACACAGCGCACTCGGAGCAGAGCG 182
Db 1862 CAGCCCTCCGGACTCGGATGACTGGAGTGTCTCCACCCCTCCCGCTCTCTGATGACCC 1921

QY 183 CTATATAAGCGACAGCCGCTCTCCGCGCCGCGCACACAGACAGATGATGCGCGCCCGAC 242
Db 1922 CCACTCTGGGCGCTCCTGCGCTGCTGTGTACCCACAGACAGACAGCAGCAGCTGCTG 1981

QY 243 CGACGGCCAGCCCGACACAACTTCTGAAAC 276
Db 1982 ACCCTGCCCTCCCTGGGCCCCCTCTCTGAAAGC 2015

RESULT 9

US-09-271-815-1
; Sequence 1, Application US/09271815
; Patent No. 6297036
; GENERAL INFORMATION:
; APPLICANT: BERGSMAN, DERK
; APPLICANT: SHABON, USMAN
; TITLE OF INVENTION: A YAK-1 RELATED SERINE/THREONINE PROTEIN KINASE-HTLAR33
; FILE REFERENCE: GH-70172-1
; CURRENT APPLICATION NUMBER: US/09/271,815
; CURRENT FILING DATE: 1999-03-18
; EARLIER APPLICATION NUMBER: 09/027,064
; EARLIER FILING DATE: 1998-02-20
; EARLIER APPLICATION NUMBER: 60/053,924
; EARLIER FILING DATE: 1997-07-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2394
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-271-815-1


```
Qy 352 CAAATGTTTGCCCAATTTGGTAAATAATTAGATGTTGAAATTAATTTTAATGACGGGC 411
|||
Db 1389 CAAATGTTTGTGCAATTTGGTAAATCTTAGATGTTGAAATTAATTTTAATGACGGGC 1448
|||
Qy 412 TCGAAGGGATTTGGTTTCGTAACCTTCGAAAATAGTCGGATGGGACAGGCGGAGGAG 471
|||
Db 1449 TCAAAGGGATTTGGTTTCGTAACCTTCGAAAATAGTCGGATGGGACAGGCGGAGGAG 1508
|||
Qy 472 AAATTTGACCGGTACCGTGTAGAGGCGGTAAATTCGAGGTTAATAATGCGACAGCAGCG 531
|||
Db 1509 AAATTTACCGGACCGTGTAGAGGCGGTAAATTCGAGGTTAATAATGCGACAGCAGCG 1568
|||
Qy 532 GTGATGACTAATAAAGAGCGCGTGAACCCCTACACCAATGCTGGAATTAATTCAGATT 591
|||
Db 1569 GTAATGACAAATAAAGACCGTCAACCCCTTATACAAATGCTGGAATTAATTCAGATT 1628
|||
Qy 592 GTGGCGCGGTCTACAGCCCGGACTTCTATGAGGACCGTGTCTGTCAGGCCCAAC 651
|||
Db 1629 GTGGGTGCGTCTACAGTCCCGAATTTCTATGAGGACCGTGTCTGTCAGGCCCAAC 1688
|||
Qy 652 CAGGAGGATCTTCATGTCAGTGGCCCGGCTTCTACCTTGTATATACCTCTGCAATGCT 711
|||
Db 1689 CAGAGGGATCTTCATGTCAGTGGCCCGGCTTCTACCTTGTATATACCTCTGCAATGCT 1748
|||
Qy 712 GGCTTTCCATATCCGGCCCGCACTGCTGACGTGTCATACCGAGGGGCTCACCTTCGAGGC 771
|||
Db 1749 GGCCTTCCGTTATCCAGCAGCCACCGCCGCGCTTACCAGGGGCGCACCTCGCAGGC 1808
|||
Qy 772 CGTGTGCGACCGGTACACACCTTACAGAGTGGGGCGGCGCCCGCCCAATCCCGCC 831
|||
Db 1809 CGCGTGCACCGGTACACACCTTACAGGCGCGGCGGCGCCCGCCCGCCCGCC 1868
|||
Qy 832 TATGCGGAGTAGTGTATCAAGAGCCAGTGTATGCAATAAATTCCTACAGGGTGTAC 891
|||
Db 1869 TACGCGGTGTGTTTACC--CAGATGATTTTATGTCGACATTTATGTTGTTAT 1926
|||
Qy 892 GCTGCATACCGCTACGCCCGCCACCCCTGCCACTGCTGCTGCTGCCTACAGTGA----- 944
|||
Db 1927 GCTGCATACCGCTACGCCCGCCAGCCTACCCCTGCGCTGCGCTGCTGCCTACAGTGA 1986
|||
Qy 945 -----CAGTTACGACGAG 958
|||
Db 1987 CAGTTGCTTCTGTTGACGACGATGAAATTTCTGTAAACACCTCTGCGAGTTACGGACGAG 2046
|||
Qy 959 TTTATGCTGCGACCCCTACACACACACACTTGTCTCCAGCCCGCCACCTACGCGTGTGTG 1018
|||
Db 2047 TTTATGCTGCGACCCCTACACACACGACCTTGTCTCCAGCCCGCCACCTACGCGTGTGTG 2106
|||
Qy 1019 CCATGAATGCTTTTGGCGCCCTTGACCGATGCGCAAGACTAGGAGCCATGCTGATGTGG 1078
|||
Db 2107 CCATGAATGCTTTTGCACCTTTGACTGATGTCGCAAGACTAGGAGCCATGCTGATGTGG 2166
|||
Qy 1079 GTCGTGCTTTTCTTCATGTCAGGCTAGTATATACAGAGGGGATACACCGTTTGTCTC 1138
|||
Db 2167 GTCGTGCTTTTCTTCATGTCAGGCTAGTATATACAGAGGGGATACACCGTTTGTCTC 2226
|||
Qy 1139 CATATTAAATGATAAAACCATTTAAACAAACAGCAAAAAAACAACAAAAA 1198
|||
Db 2227 CATACTAATGCAAAACCATTA----- 2250
|||
Qy 1199 CAACCTTCCAATGTGGGAGAGGAAGCTTTCCAGGCGCGAGTGTTCGACACATGCA 1258
|||
Db 2251 -AACCTTCCAATGTGGGAGAAAGGAGCTTTCCAGGCGCTGAGTATTGCAATACATGCA 2309
|||
Qy 1259 GTAGACATCACTTTAGCAACTCAAGAAACAAACAAAAAATAA 1318
|||
Db 2310 GTAGTACATCACTTTAGCAATTTTAAAAAATAAATAAATAAATAA 2369
|||
```

RESULT 2

US-09-176-657-5

; Sequence 5, Application US/09176657

```
; Patent No. 6020164
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Tang, Y. Tom
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Lu, Aina
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: HUMAN RNA BINDING PROTEINS
; FILE REFERENCE: P2-0611 US
; CURRENT APPLICATION NUMBER: US/09/176,657
; CURRENT FILING DATE: 1998-10-21
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 1506
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 1250374
US-09-176-657-5
```

```
Query Match 20.6%; Score 275.6; DB 3; Length 1506;
Best Local Similarity 60.2%; Pred No. 8.5e-69;
Matches 601; Conservative 0; Mismatches 354; Indels 43; Gaps 7;
```

```
Qy 222 GACAGATGATGCGCCCGCCGACCGACGCGCCAGC-----CCAGACACAACTTCTGAAAAAC 278
|||
Db 452 GACAGAAGGTGAGACACAGACAGACGCGCCAGCAGTCAAGACACAAAGTAGTGAATTC 511
|||
Qy 279 AGAAAAAAGTCCAGGCCCAAGCGGCTGCATGTGTCAACATCCCTTCGGTTCGCGGA 338
|||
Db 512 AGAGAGTAAATCTACCCCGAAACGCTGTCATGCTCTAAATATCTTCCGCTTCGCGGA 571
|||
Qy 339 TCCAGACTCCGACAAATGTTTGCCCAATTTGGTAAAAATATTAGATGTTGAAATTAATTT 398
|||
Db 572 CCCTGACTCCGGCAGATGTTTGGGCAGTTTGGCAGAAATCCTAGATGTAGAAAAATCTT 631
|||
Qy 399 TAATGAGCGGGCTCGAAGGATTTGGTTTCGTAACCTTCGAAAAATAGTGCAGTGGCGA 458
|||
Db 632 TAATGAACGTGGCTCTAAGGGATTCGGGTTTCGTAACCTTCGAGAATAGTGTGATGAGA 691
|||
Qy 459 CAGGCGAGGAGGAAAAATTGACGGTACCGGTAGAGGCGCGTAAAAATCGAGGTTAATAA 518
|||
Db 692 CAGGCGCAGGAGAAAAATTACACGGCACCGTGTAGAGGCGCGTAAAAATCGAGGTTAATAA 751
|||
Qy 519 TGCACACGACGCGTGTGACTAATAAAGCGCGTGAACCCCTACACCAATGGCTGCGAA 578
|||
Db 752 TGCTACAGCAGTGTAAATGACCAATAAGAGATGGTCAACCATATGCAATGGTTGAA 811
|||
Qy 579 ATTAATTCAGTTCGTGGCGCGGCTTACAGCCCGGACTTCTATGAG-----GCA 628
|||
Db 812 ATTAAGCCCACTAGTTGGAGCTGTATATGTCGCGAGTTATATGACGATTCAGCTTCA 871
|||
Qy 629 CGGTGCTGTTGTCAGGCCCAACAGAGGGATCTTCCATGTA--CAGTGGCCCCAGTTC 686
|||
Db 872 AGCAGATGTCCTTAGGCAATGATGACAGAGTGCCCGCTATCAGAGAGAGGGGATCAAA 931
|||
Qy 687 ACTGTATATCTTCTGCAATGCTGGCTTTCCATA---TCGGCGCGCACTGCTGTCAGC 743
|||
Db 932 CACTTACATTCCTTTAATCATCTTCTGGCTTCCCTTACCTACTGACGACACACGCGAGC 991
|||
Qy 744 TGCATACCGAGGGGCTCACCTTCGAGGCGCGTGGTCGACCGGTATCAACACCTTCAGAGC 803
|||
Db 992 CGCTTTCAGAGAGGCCCAATTTGAGGGGCGAGGGCGGACAGTATATGTCAGTCCGA-- 1049
|||
Qy 804 TGGGCGCCCCACCCCAATCCCGGCTATGGCGGAGTAGTGTATCAAGAGCCAGTGTGA 863
|||
Db 1050 -GCGGTACTCTCAACAGCCATCCCGGCTATCCAGGGGTGGATATGACAGCCTACAG---- 1104
|||
Qy 864 TGGCAATAAATGCTACAGGGTGGTTACGCTGCATACCGCTACGCCCGCCAGCCCGCTGC 923
|||
Db 1105 -----ATATGACAGCCTGCTACTGCAACCGCAGCCACCGCTGCTGTCAGC 1149
|||
```


THIS PAGE BLANK (USPTO)

/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-E0-bv-c-04-0-UI"
/clone_lib="UI-R-E0"
/dev_stage="embryonic"
/lab_host="DHI08 (Life Technologies)"
/notes="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site 1: NotI; Site 2: EcoRI; This library
consists of a mixture of individually tagged normalized
libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dr track which allows
identification of the library of origin of a clone within
the mixture."

BASE COUNT 119 a 156 c 100 g 85 t

Query Match 28.6%; Score 383.2; DB 9; Length 460;
Best Local Similarity 98.0%; Pred. No. 6e-59;
Matches 451; Conservative 0; Mismatches 3; Indels 6; Gaps 6;

QY 2 CGCGCGCCCTGACACATGCTCAG-CTTATGCTCAGCGCAGTTCGCTCCACCCGAGA 60

DB 1 CGCGCGCCCTGACACATGCTCAG-CTTATGCTCAGCGCAGTTCGCTCCACCCGAGA 60

QY 61 ATGGCAT-CCTGCGAATACACGGCCCTCATCCCTCCCGCCGAG-ACACCGGCC 118

DB 61 ATGGCATCCCTGCGAATACACGGCCCTCATCCCTCCCGCCGAGTACACCGGCC 120

QY 119 AGCCCATGTCCTCG-CCACACATTAACCTTG-ATCTCTTACACAGCGGACTCGGAGC 176

DB 121 AGACCACTGTCCCGACACACATTAACCTTGATCTCTTACACAGCGGACTCGGAGC 180

QY 177 AGAGCGCTTATACAGCGCAG-CCGTCTCCGCGACCGCCACACAGACATGATGCGG 235

DB 181 AGAGCGCTGATACAGCGCAGACCGTCTCCGCGACCGCCACACAGACATGATGCGG 240

QY 236 CCCCGACCGAGCGGCGCCCGACACAACTTCTGAAAAACAGAAACAGTCCCGAGC 295

DB 241 CCCCGACCGAGCGGCGCCCGACACAACTTCTGAAAAACAGAAACAGTCCCGAGC 300

QY 296 CCAAGCGGCTGATGTGCAACATCCCTTCGGTTCGGGATCCAGACCTCCGACAA 355

DB 301 CCAAGCGGCTGATGTGCAACATCCCTTCGGTTCGGGATCCAGACCTCCGACAA 360

QY 356 TGTGTGGCAATTTGGTAAATATTAGATGTGAATATTATTTAAATGAGCGGGCTCGA 415

DB 361 TGTGTGGCAATTTGGTAAATATTAGATGTGAATATTATTTAAATGAGCGGGCTCGA 420

QY 416 AGGATTTGGTTTCGTAACCTTCGAAAAATAGTGGCGATGC 455

DB 421 AGGATTTGGTTTCGTAACCTTCGAAAAATAGTGGCGATGC 460

RESULT 14
AV665025
LOCUS AV665025 549 bp mRNA linear EST 28-NOV-2001
DEFINITION AV665025 Bos taurus brain fetus Bos taurus CDNA clone E1BR045H01
5', mRNA sequence.
ACCESSION AV665025
VERSION AV665025.1 GI:9924055
KEYWORDS EST.
SOURCE cow.

ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 549)
Takasuga, A., Hirotsune, S., Itoh, R., Jitohzono, A., Suzuki, H., Aso, H.
and Sugimoto, Y.
REFERENCE
AUTHORS

TITLE
Establishment of a high throughput EST sequencing system using
poly(A) tail-removed cDNA libraries and determination of 36,000
bovine ESTs

JOURNAL Nucleic Acids Res. 29 (22), E108 (2001)
MEDLINE 21570554
COMMENT Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@coo.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.

FEATURES
source

1..549
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="E1BR045H01"
/clone_lib="Bos taurus brain fetus"
/tissue_type="brain"
/dev_stage="fetus"
/lab_host="DHI08"
/notes="Vector: pZLI; Site_1: SalI; Site_2: NotI; Poly A
was deleted from a NotI site"
151 a 151 c 141 g 104 t 2 others

Query Match 28.3%; Score 379.8; DB 10; Length 549;
Best Local Similarity 84.3%; Pred. No. 2.3e-58;
Matches 463; Conservative 0; Mismatches 79; Indels 7; Gaps 3;

QY 84 CCCTCATCCCCATCCCGCCGACAGACCGGCCAGCCCACTGTCCCG-CCACACATT 142

DB 1 CCAGCGCTCCGCGCCAGCTCCGAGTACACTGGCCAGACCCGTCGCCGAGCACAGTT 60

QY 143 AAACCTTG-ATCTCTCTACACAGACGCACTCGGAGCAGAGCGCTTATACAGCGCACAC 199

DB 61 AAACCTTGATCCCTCCCGCGAGTCGCACTCGGAGCAGAGCGCGGAGCAGCTAGCGCGC 120

QY 200 ---CCGTCTCCGGCACCAGCCACACAGACATGATGCGGCCCGCCAGCCAGCGGCCCGCC 256

DB 121 ACACCGTCTCCGGCACCAGCCACACAGACATGATGCTCGGCACCCACCGATGGCCAGCCCC 180

QY 257 AGACACAACTTCTGAAACACACAGAAACAAAGTCCAGGCCCAAGCGCTGATGTGCTCA 316

DB 181 AGACACAGCTTCTGAAACACAGAAACAAAGTCCAGGCCCAAGCGCTGATGTGCTCA 240

QY 317 ACATCCCTTCCGGTTCGGGATCCAGACCTCCGACAAATGTTTGGCCAAATTTGGTAAAA 376

DB 241 ACATCCCTTCCGGTTCGGGATCCGGACCTCCGACAAATGTTTGGTCAATTTGGTAAAA 300

QY 377 TATTAGATGTTGAAATTTATTTTAAATGAGCGGGCTCGAAGGGATTTGGTTTCGTAACCT 436

DB 301 TCTTAGATGTTGAAATTTATTTTAAATGAGCGAGGCTCAAAGGGATTTGGTTTCGTAACCT 360

QY 437 TCGAAATAGTTCGGATCGGACAGGCGGAGGAGAAATTCACCGGTACCGTGGTAGAGG 496

DB 361 TCGAAATAGTTCGGATCGGACAGGCGGAGGAGAAATTCACCGGACCGTGGTAGAGG 420

QY 497 GCCGTAAAAATCAGGTTAAATATCGGACAGCACGCGTGTGATGACTAATAAAGGCGGTGA 556

DB 421 GCCGTAAAAATCAGGTTAAATATCGGACAGCACGCGTGTGATGACTAATAAAGGCGGTGA 480

QY 557 ACCCTACACCAATGGCTGGAAATTAATCCAGTTGTGGGCGCGGTCTACAGCCCGGACT 616

DB 481 ACCCTTATACAAATGGCTGGAAATTAATCCAGTTGTGGGTCAGNGTACAGTCTCTGAAT 540

QY 617 TCTATGAG 625

DB 541 TCTATGAG 549

RESULT 15
A1193843/c
LOCUS A1193843 451 bp mRNA linear EST 29-OCT-1998

```
QY 1013 TTGGTGCATGAATGCTTTTGGCCCTTGACCGATGCGCAAGACTAGAGCCATGCTGATG 1072
Db 298 TTGGTGCATGAATGCTTTTGGCACTTTGACTGATGCGCAAGACTAGAGCCATGCTGATG 357
QY 1073 ATGGGTCTCGTCTTTCTTCTTCAATGCGAGCTAGTATATACCAAGGGGATACAACCGTT 1132
Db 358 ATGGGTCTCGTCTTTCTTCTTCAATGCGAGCTAGTATATACCAAGGGGATACAACCGTT 417
QY 1133 TTGCTCCATATTAAATGATATAAACCATTAAACAAACGCAAAACAAACAAACAA 1192
Db 418 TTGCTCCATATTAAATGATATAAACCATTAAACAAACGCAAAACAAACAAACAA 447
QY 1193 AAAAACCAACTTCCCAATGTTGGGAGAGAGAAAGCTTTCCGAGCCCGAGTGTTCGACA 1252
Db 448 -----AACCTTCCAATGTGGGAGAGAAAGCTTTCCGAGCCCGAGTATTGCAATA 500
QY 1253 CATGCGTAGGACATCATCTTTAGCAACTCAAGNACACGCAAAAA 1299
Db 501 CATGCGTAGTATCATCTTTTAGCAACTCTAAAAAATAAAAAA 547

RESULT 12
AV729198 773 bp mRNA linear EST 17-OCT-2000
LOCUS AV729198 HTC Homo sapiens cDNA clone HTCBCD09 5', mRNA sequence.
ACCESSION AV729198
VERSION AV729198.1 GI:10838619
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 773)
Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu
, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R.,
Chen, J., Chen, Z. and Han, Z.
Homo sapiens cDNA HTC clones
Unpublished (2000)
Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
1. .773
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTCBCD09"
/clone_lib="HTC"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 184 a 232 c 185 g 170 t 2 others
ORIGIN
Query Match 28.7%; Score 384.2; DB 10; Length 773;
Best Local Similarity 82.1%; Pred. No. 3.3e-59;
Matches 482; Conservative 0; Mismatches 65; Indels 40; Gaps 2;
QY 713 GCTTTTCATATCGGCGCGCACTGCTGCAGCTGCATACCGAGGGGTCACTTCGAGGCC 772
Db 1 GCTTCCGTATCCAGCAGCACCGCTCGCGCGCTTACCGAGGGGCGCACCTTCGAGGCC 60
QY 773 GTGGTCGACCGTGTACACACCTTCAGAGCTCGGGGCGCCCAACCCCAATCCCGGCT 832
Db 61 GCGGTGCGACCGTGTACACACCTTCAGGCGCGCGGCGCCCGCCCGGCGCT 120
```



```

Db 301 CCCAAGCGGTGCATGTCTCAACATCCCCCTTCGGTTCGGGATCCAGACCTCCGACAA 360
Qy 355 ATGTTTGGCCAAATTTGTAATAATTAGATGTTGAAATTAATTTTAATAGCGGGGCTCG 414
Db 361 ATGTTTGGCCAAATTTGTAATAATTAGATGTTGAAATTAATTTTAATAGCGGGGCTCG 420
Qy 415 AAGGATTTGGTTTCGTAACCTTTCGAAAAATAGTCGGATCGGACAGGGCGAGGAGAAA 474
Db 421 AAGGATTTGGTTTCGTAACCTTTCGAAAAATAGTCGGATCGGACAGGGCGAGGAGAAA 480
Qy 475 TTGCACGGTACCGT 488
Db 481 TTGCACGGTACCGT 494

RESULT 7
AI291784/c
LOCUS AI291784 479 bp mRNA linear EST 29-JAN-1999
DEFINITION qm7508.x1 Soares_placenta 8to9weeks 2NbHP8to9W Homo sapiens CDNA
clone IMAGE:1894551 3' similar to TR:Q10572 Q10572 44.4 KD
RNA-BINDING PROTEIN IN FOX-1 LOCUS. ;, mRNA sequence.
ACCESSION AI291784
VERSION AI291784.1 GI:3934558
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 479)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 863 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 452.
FEATURES
source
1..479
/organism="Homo sapiens"
/db xref="taxon:9606"
/clone="IMAGE:1894551"
/clone_lib="Soares_placenta 8to9weeks 2NbHP8to9W"
/dev_stage="two placentae: One from 8 weeks and another
from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: placenta; Vector: p7T73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGATTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p7T73 vector
(Pharmacia). Library constructed by Bento Soares and
M. Patricia Bonaldo."
BASE COUNT 112 a 118 c 115 g 134 t
ORIGIN
Query Match 30.9%; Score 413.4; DB 9; Length 479;
Best Local Similarity 91.4%; Pred. No. 2.3e-64;
Matches 438; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 268 TCTGAAAAACACAGAAAAAAGTCCAGCCCGGCTGTCATGTGTCCAAATCCCTTC 327
Db 479 TCTGAAAAACACAGAAAAAAGTCTCAGCCAGCGCTGTCATGTCTCCAAATCCCTTC 420
Qy 328 CGGTTCCGGATCCAGACCTCCGACAAATGTTTGGCCAAATTTGGTAAATATTAGATGTT 387
Db 419 AGGTTCCGGATCCCGGACCTCAGACAAATGTTTGGTCAATTTGGTAAATATTAGATGTT 360

```

```

Qy 388 GAAATTATTTTAAATCAGCGGGGCTCGAAGGGATTTGGTTTCGTAACCTTTGAAATAGT 447
Db 359 GAAATTATTTTAAATCAGCGAGGCTCAAGGGGATTTGGTTTCGTAACCTTTGAAATAGT 300
Qy 448 GCCGATTCGCGNACAGGCGAGGGAGAAATTCACCGGTACCGTGTAGAGGGCCGCTAAATC 507
Db 299 GCCGATTCGCGNACAGGCGAGGGAGAAATTCACCGCACCGTGTAGAGGGCCGCTAAATC 240
Qy 508 GAGGTTAATTAATCGCACAGCACCGGTGATGACTAATAAAAAAGCGCGTGAACCCCTACACC 567
Db 239 GAGGTTAATTAATCGCACAGCACCGGTGATGACTAATAAAAAAGCGCTGAACCCCTTATACA 180
Qy 568 AATGGCTGGGAATTAATCCAGTTGTGGCGCGTCTACAGCCCGCACTTCTATGCAAGGC 627
Db 179 AATGGCTGGGAATTAATCCAGTTGTGGGTGCAGTCTACAGTCCCGAATTTCTATGCAAGC 120
Qy 628 ACGGTGCTGTGTGCCAGGCAACACGAGGGGATCTTCCATGTACAGTGGCCCGCAGTTCA 687
Db 119 ACGGTGCTGTGTGCCAGGCAACACGAGGGGATCTTCCATGTACAGTGGCCCGCAGTTCA 60
Qy 688 CTTGTATATATCTTCTGCAATGCGTGTTCCTTCCATATCCGGCGCCCACTGCTGCAAGTCC 746
Db 59 CTTGTATATATCTTCTGCAATGCCAGGCTTCACGTATCCAGCAGCCACCGCGCGCGCGC 1

RESULT 8
BM808707
LOCUS BM808707 1020 bp mRNA linear EST 05-MAR-2002
DEFINITION AGENCOURT 6618011 NIH_MGC_124 Homo sapiens CDNA clone IMAGE:5734270
5', mRNA sequence.
ACCESSION BM808707
VERSION BM808707.1 GI:19125530
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1020)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Invitrogen
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L2AMI2739 row: e column: 23
High quality sequence stop: 723.
FEATURES
source
1..1020
/organism="Homo sapiens"
/db xref="taxon:9606"
/clone="IMAGE:5734270"
/clone_lib="NIH MGC 124"
/tissue_type="hippocampus"
/lab_host="DH10B"
/notes="Organ: Brain; Vector: pCMV-SPORT6; Site 1: EcoRV
(destroyed); Site 2: NotI; RNA source male hippocampus,
age 27. Library is oligo-dT primed and directionally
cloned (EcoRV site is destroyed upon cloning). Average
insert size 1.4 kb, insert size range 0.9-4 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 012."
BASE COUNT 274 a 279 c 275 g 189 t 3 others
ORIGIN
Query Match 30.0%; Score 401.6; DB 14; Length 1020;
Best Local Similarity 86.5%; Pred. No. 2.3e-62;

```


ORIGIN

QY 302 GGCTGATGTGTCACATCCCTTCCTCGGTTCGGGATCCAGACCTCCGACAAATGTTTG 361
 Db 301 GGCTGATGTGTCACATCCCTTCCTCGGTTCGGGATCCAGACCTCCGACAAATGTTTG 360
 QY 362 GCCAATTTGGTAAATATATAGATGTTGAAATATATTTTAAATGAGCGGGCTCGAAGGAT 421
 Db 361 GCCAATTTGGTAAATATATAGATGTTGAAATATATTTTAAATGAGCGGGCTCGAAGGAT 420
 QY 422 TTGGTTTCGTAACTTTTCGAAATAGTCGGGATCGGACAGAGCGGCGAGGAGAAATTTGCACG 481
 Db 421 TTGGTTTCGTAACTTTTCGAAATAGTCGGGATCGGACAGAGCGGCGAGGAGAAATTTGCACG 480
 QY 482 GTACCGTGTAGAGCGGCGTAAATTCGAGGTTTAATATCGACAGCACGCGTGATGACTA 541
 Db 481 GTACCGTGTAGAGCGGCGTAAATTCGAGGTTTAATATCGACAGCACGCGTGATGACTA 540
 QY 542 AT 543
 Db 541 AT 542

RESULT 4
 A1885361/c
 LOCUS
 DEFINITION w192h03.x1 NCI CGAP Brn25 Homo sapiens cDNA clone IMAGE:2432405 3' similar to TR:O43251 O43251 HYPOTHETICAL 39.5 KD PROTEIN. ;, mRNA sequence.
 A1885361
 A1885361.1 GI:5590525
 VERSION EST.
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 587)
 NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGA), Tumor Gene Index
 Unpublished (1998)
 JOURNAL
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 728 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 459.
 Location/Qualifiers
 1..587
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:2432405"
 /tissue_type="anaplastic oligodendroglioma"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCGAAGTCGAGCGCGCATAGGTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo."
 121 a 143 c 156 g 167 t

BASE COUNT

Query Match 34.6%; Score 464.2; DB 9; Length 587;
 Best Local Similarity 90.1%; Pred. No. 1.8e-73;
 Matches 520; Conservative 0; Mismatches 53; Indels 4; Gaps 2;

QY 141 TTAACCTTGATCTCTCTACACAGACGCACTCGGAGCAGAGC---GCTTATACAGCCGAC 197
 Db 577 TAAACCTGTATACCTCCGCGCCAGACGCACTCCGAGCAGAGCCCGCGGACACAGCGCTC 518
 QY 198 AG-CCGTCCTCGGACCGCCACACAGACATGATGCGGCCCGCCACGACGCGCCAGCCCC 256
 Db 517 AGACCGTCTCTGGCACCAGCCACACAGACATGACGACGACCGACCGATGGCCAGCCCC 458
 QY 257 AGACACAACCTTCTGAAAAACACAGAAACAAAGTCCAGAGCCCAAGCGCTGCATGTGTCCA 316
 Db 457 AGACACAACCTTCTGAAAAACACAGGAAACAAAGTCTCAGCCCCAAGCGCTGCATGTCTCCA 398
 QY 317 ACATCCCTTCCGTTCCGGATCCAGACCTCCGACAAATGTTTGGCCCAATTTGGTAAAAA 376
 Db 397 ATATCCCTTTCAGGTTCGGGATCCGACCTCCGACCAATGTTTGGTCAATTTGGTAAAAA 338
 QY 377 TATTAGATGTTGAAATTTATTTTAAATGAGCGGGCTCGAAGGGATTTGGTTTCGTAACTT 436
 Db 337 TCTTAGATGTTGAAATTTATTTTAAATGAGCGAGCTCAAGGGATTTGGTTTCGTAACTT 278
 QY 437 TCGAAAAATAGTCGGATCGGACAGCGAGCGAGGAGAAATTCACCGGTACCGTGTGTAGAG 496
 Db 277 TCGAAAAATAGTCGGATCGGACAGCGAGCGAGGAGAAATTCACCGGTGTGTAGAGG 218
 QY 497 GCCGTAATAATCGAGTTAATAATCGGACAGCACCGCGTGTAGTACTAATAAAGAGCCGTGA 556
 Db 217 GCCGTAATAATCGAGTTAATAATCGGACAGCACCGCTGTAAAGACAAATAAAGAGCCGTGA 158
 QY 557 ACCCTTACACCAATGCGTGGAAATTAATCCAGTTGTGGCGCGGTCTACAGCCCGGACT 616
 Db 157 ACCCTTATACAAATGCGTGGAAATTAATCCAGTTGTGGGTGAGTCTACAGTCCCGGAT 98
 QY 617 TCTATGAGGACGCGTGTGTGTGCGAGGCAACACGAGGAGGATCTTCCATGTACAGTG 676
 Db 97 TCTATGAGGACGCGTGTGTGTGCGAGGCAACACGAGGAGGATCTTCCATGTACAGTG 38
 QY 677 GCCCAGTTTCACTTGTATATACTTCTGCAATGCCTGG 713
 Db 37 CCCCCAGTTTCACTTGTATATATCTTCTACAATGCCAGG 1

RESULT 5
 BF549922/c
 LOCUS
 DEFINITION BF549922 533 bp mRNA linear EST 12-DEC-2000
 UI-R-E0-bv-C-04-0-UI.r1 UI-R-E0 Rattus norvegicus cDNA clone
 UI-R-E0-bv-C-04-0-UI 5', mRNA sequence.
 ACCESSION BF549922
 VERSION BF549922.1 GI:11659652
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 1 (bases 1 to 533)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene discovery
 Genome Res. 6 (9), 791-806 (1996)
 JOURNAL
 MEDLINE
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu

Contact: nisc_mgc@nhgri.nih.gov
 Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
 Maduro, Q.L., Masiello, C., Maekeri, B., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
 Tsurgelon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 79 Row: j Column: 17
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 8922072
 This clone has the following problem: frame shifted.

FEATURES

source

1. 973
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5734270"
 /tissue_type="Brain, hippocampus"
 /clone_lib="NIH MGC_124"
 /lab_hosts="DHIOB"
 /notes="Vector: pCMV-SPORT6"

BASE COUNT 265 a 270 c 256 g 182 t
 ORIGIN

Query Match 39.5%; Score 529.2; DB 11; Length 973;
 Best Local Similarity 88.9%; Pred. No. 3.7e-85;
 Matches 663; Conservative 0; Mismatches 73; Indels 10; Gaps 8;
 QY 1 GCGCGCCCTGTACACAAATGGCTCAG-CTTATGCTCAGCGCAGTTTCGTCTCCACCCCGAG 59
 DB 228 GCCGCTGCCCTGTACACAAATGGCTCAGCCTTACGCTTCGGCCAGTTTGTCTCCCGCAG 287
 QY 60 AATGGCATCTT-GCAGAAATACACGGCCCTCATCCCATCCCGCCGCGAGAG-ACACCGCG 117
 DB 288 AACGGTATCCCGCGGAATACACGGCCCTCATCCACCCCGCGCAGAGTACACAGGC 347
 QY 118 CAGCCCACTGTCCCG-CCACACATTAACCTTG-ATCGTCTACACAGCGCCTCGAG 175
 DB 348 CAGACCAAGTTCCTCCGAGCAGACATTAACCTGTACCTCTCCGCCAGCGCACTCCGAG 407
 QY 176 CAGAGC--GCTTATACAAGGGCAGAG-CCGTCTCCGCGACCGCCACACAGACAGATGAT 231
 DB 408 CAGAGCCGGGGGACACAGAGCGCTCAGACCGCTCTCTGGCAGCCGACACAGACAGATGAC 467
 QY 232 GCGCCCGGACCGGACCGCCAGCCAGACACAACTTTCTGAAAAACACAGAAAAACAAGTCC 291
 DB 468 GCAGCAGCGGATGGCAGCGCCAGACACAACTTTCTGAAAAACACGGAACAAGTCT 527
 QY 292 CAGCCCAAGCGGCTGCATGTGTCAACATCCCTTCGGTTCGGGATCCAGACTCCGA 351
 DB 528 CAGCCCAAGCGGCTGCATGTGTCAACATCCCTTCAGGTTCCGGGATCCGGAGCTCAGA 587
 QY 352 CAATGTTTGGCCAAATTTGGTAAATATATAGATGTTGAAATATTTTAAATGAGCGGGC 411
 DB 588 CAATGTTTGGTCAATTTGGTAAATATTTAGATGTTGAAATATTTT-AAATGAGGAGGC 646
 QY 412 TCGAAGGATTTGGTTTCGTAACCTTTTCGAAAAATAGTCGGATCGGAGCGGAGGAGG 471
 DB 647 TCAAGGATTTGGTTTCGTAACCTTTTCGAAAAATAGTCGGATCGGAGCGGAGGAG 706
 QY 472 AAATTGACCGGTACCGGTAGAGGGCGGTAAATATCGAGGTTAATATGCGACAGCAGCG 531
 DB 707 AAATTACAGCGCACCGGTAGAGGGCGGTAAATATCGAGGTTAATATGCGACAGCAGCT 766
 QY 532 GTGATGACTAATAAAGGCGGTGAACCCCTACACCAATGCTGGAATTAATTCAGATT 591
 DB 767 GTAATGACAAATAAAGACCGGTCAACCCCTTATCAATGCTGGAATTAATTCAGATT 826
 QY 592 GTGGCGCGGTCTACAGCCCGCACTTCTATGACGCGCGGTGCTGTGTGTGCGAGGCCAAC 651

Db 827 GTGGTGCACTGTACAGTCCCGAAATTTCTATGAGGACGGTCTGTGTGCCAGGCCAAC 886
 QY 652 CAGAGGGATTTCCCATGTACAGTGGCCCCAGTTCCTTGTATATATCTTCTGCAATGCGCT 711
 DB 887 CAGAGGGATTTCCCATGTACAGTGGCCCCAGTTCCTTGTATATATCTTCTGCAATGCCA 946
 QY 712 GGTTTTCCATATCCGGCCGCACTGC 737
 DB 947 GGCTTCCGTATCCAGCAGCCACCGC 972
 RESULT 3
 AI101278
 LOCUS
 DEFINITION EST210567 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
 RBRBK70 3' end, mRNA sequence.
 ACCESSION AI101278 GI:3704621
 VERSION AI101278
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 542)
 AUTHORS Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,
 Kerlavage, A.R. and Adams, M.D.
 TITLE Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
 Gene Index
 JOURNAL Unpublished (1998)
 COMMENT Contact: Lee, NH
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: M13-21.
 FEATURES
 Location/Qualifiers
 1. 542
 /organism="Rattus sp."
 /db_xref="ATCC (inhost):2022664"
 /db_xref="taxon:10118"
 /clone="RBRBK70"
 /clone_lib="Normalized rat brain, Bento Soares"
 /note="Organ: brain; Vector: pT7T3pac; Site_1: EcoRI;
 Site_2: NotI"
 BASE COUNT 146 a 168 c 128 g 100 t
 ORIGIN
 Query Match 34.7%; Score 465.2; DB 9; Length 542;
 Best Local Similarity 98.3%; Pred. No. 1.2e-73;
 Matches 533; Conservative 0; Mismatches 3; Indels 6; Gaps 6;
 QY 8 CCCTTGACAAATGGCTCAG-CTTATGCTCAGCGAGTTTCGTCCACCCAGAAATGGCA 66
 DB 1 CCCTTGACAAATGGCTCAGCCTTATGCTTCAGCGAGTTTCGTCCACCCAGAAATGGCA 60
 QY 67 T-CCTGAGAAATACAGCGCCCTCATCCCATCCCGCCGAGAG-ACACCGGCGAGCCCA 124
 DB 61 TCCTGAGAAATACAGCGCCCTCATCCCATCCCGCCGAGAGTACACCGGCGAGACCA 120
 QY 125 CTGTCCCG-CCACACATTAACCTTG-ATCCTCTACACAGCGCACTCGGAGCAGAGCG 182
 DB 121 CTGTCCCGACCAACATTAACCTGTATCTCTACAGAGCGCACTCGGAGCAGAGCG 180
 QY 183 CTTATACAAGCGCACAG-CCGTCTCCGGCACCCGCCACACAGACAGATGATGCCGCCCGCA 241
 DB 181 CTGATACAAGCGCACAGCGCTCTCCGGCACCCGCCACACAGACAGATGATGCCGCCCGCA 240
 QY 242 CCGACGGCAGCGCCAGACACAACTTCTGAAAAACACAGAAAAACAAGTCCAGCCCAAGC 301
 DB 241 CCGACGGCAGCGCCAGACACAACTTCTGAAAAACACAGAAAAACAAGTCCAGCCCAAGC 300

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, J., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 32 Row: 1 Column: 9
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 8922072
 This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers
 1..4039
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4815500"
 /tissue type="Brain, hippocampus"
 /clone lib="NIH MGC_95"
 /lab host="DH10B"
 /note="vector: pBluescript"
 BASE COUNT 1201 a 858 c 876 g 1104 t
 ORIGIN

Query Match 68.0%; Score 911.2; DB 11; Length 4039;
 Best Local Similarity 85.9%; Pred. No. 1.2e-153;
 Matches 1144; Conservative 0; Mismatches 138; Indels 50; Gaps 10;

```

Qy 1 GCGGCCGCCCTGACAAATGGCTCAG-CTTATGCTCTCAGCGCAGTTTCGCTCCACCCCGAG 59
Db 280 GCACTGCCCCCTGACAAATGGCTCAGCGCTTACGCTTCGGCCAGTTTGTCTCCCGCCGAG 339

Qy 60 AATGCGATCTT-GCAGAATACAGCGCCCTCATCCCATCCGCGCCAGAG-ACACCGC 117
Db 340 AACGGTATCCCGCGGAATACACGCGCCCTCATCCCATCCGCGCCAGAGTACACAGCG 399

Qy 118 CAGCCCACTGTCCTCCG-CCACACATTAACCTTG-ATCTCTCTACACAGCAGCCTCGAG 175
Db 400 CAGACACAGGTTCCCGAGCACAATTAACCTTGATCCCTCCGCGCCAGCAGCTCCGAG 459

Qy 176 CAGAGC---GCTTATACAGCGCACAG-CCGTCTCCGCGCACCGCCACACAGACATGAT 231
Db 460 CAGAGCCCGCGGACACAGCGCTCAGACCGGTATCTGGCAGCCGCCACACAGACATGAC 519

Qy 232 GCGGCCGACCGAGCGCCAGCCAGACACAACTTCTGTAACACACAGAAACAGTCC 291
Db 520 GCAGCAGCGCGGATGGCCAGCGCCAGACACAACTTCTGTAACACACAGAAACAGTCT 579

Qy 292 CAGCCCAAGCGCTGATGTGTCACATCTCCCTTCGGTTCGGGATCCAGACCTCCGA 351
Db 580 CAGCCCAAGCGCTGATGTGTCACATCTCCCTTCGGTTCGGGATCCAGACCTCCGA 639

Qy 352 CAAATGTTTGGCCAAATTTGGTAAATATTTAGATGTTGAAATATTTTAAATGAGCGGGC 411
Db 640 CAAATGTTTGGTCAATTTGGTAAATCTTTAGATGTTGAAATATTTTAAATGAGCGAGGC 699

Qy 412 TCGAAGGATTTGGTTTGTCTTAATCTTCGAAATATGTCGGATCGGACAGGCGAGGAG 471
Db 700 TCAAGGATTTGGTTTGTCTTAATCTTCGAAATATGTCGGATCGGACAGGCGAGGAG 759

Qy 472 AAATTCACGGTACCGTGTAGAGGCGCGTAAATTCGAGGTTAAATTCGACACAGCAGC 531
Db 760 AAATTCACGGTACCGTGTAGAGGCGCGTAAATTCGAGGTTAAATTCGACACAGCAGC 819

Qy 532 GTGATGACTTAATAAAGGCGGTGAACCCCTACACCAATGGCTGGAAATTAATCCAGTT 591
Db 820 GTAATGACAAATAAAGAGCCGTCAACCTTATACAAATGGCTGGAAATTAATCCAGTT 879

Qy 592 GTGGGCGCGTCTACAGCCCGACTCTATGAGGACCGGTCTGTTGCGAGGCGCAAC 651
Db 880 GTGGGTCAGTCTACAGTCCCGAATTTCTATGAGGACCGGTCTGTTGCGAGGCGCAAC 939

Qy 652 CAGGAGGATCTTCATGTACAGTGGCCCGCCAGTTCACTTGTATATCTCTCAATGCT 711
Db 940 CAGGAGGATCTTCATGTACAGTGGCCCGCCAGTTCACTTGTATATCTCTCAATGCTCA 999

```

```

Qy 712 GCGTTTCATATCCGGCGCGGCGCTGCTGAGCTGCATACCGAGGGGCTCACCTTCGAGGC 771
Db 1000 GCGTTTCATATCCAGCAGCAGCCGCGCGGCGCTTACCGAGGGGCGCACCTTCGAGGC 1059

Qy 772 CGTGTGCGACCGGTGTAACACACCTTCAGAGCTGCGCGGCGCCACCCCAATCCCGGCC 831
Db 1060 CGCGTTCGACCGGTGTAACACACCTTCAGGGCGCGGCGGCGCCCGCCGATCCCGGCC 1119

Qy 832 TATGCGGAGTGTATCAAGAGCCAGTGTATGGCAATAAATTTGTCTACAGGGTGGTTAC 891
Db 1120 TACGGCGGTGTTGTTTACCAGGATGGATTTATGG---TGCAGACATTTATGTTGGTTAT 1176

Qy 892 GCTGCATACCGCTACGCGCCAGCCACCCCTGCGCACTGCTGCTCTCTACAGTGACAGTAC 951
Db 1177 GCTGCATACCGCTACGCGCCAGCCCTACCCCTGCGCACTGCTGCTCTACAGTGACAGTAC 1236

Qy 952 GGAGAGTTTATGCTGCGACCCCTTACCCACACACTTGTCTCAGCGCCCGCCACCTACGCGC 1011
Db 1237 GGAGAGTTTATGCTGCGACCCCTTACCCACACACTTGTCTCAGCGCCCGCCACCTACGCGC 1295

Qy 1012 GTTGTGCTCAATGCTTTTGGCGCCCTTGACCGATGCCAAGACTTAGGAGCCATGCTGTAT 1071
Db 1296 GTTGTGCTCAATGCTTTTGGCACCTTTTGACCTGATGCCAAGACTTAGGAGCCATGCTGTAT 1355

Qy 1072 GATGTGGTCTCGTCTTCTTCTTCTGATGCGAGCTAGTATATACCAAGGGGATACACCGT 1131
Db 1356 GATGTGGTCTCGTCTTCTTCTTCTGATGCGAGCTAGTATATACCAAGGGGATACACCGT 1415

Qy 1132 TTTGCTCCATATTAATGATAAAACCAATTAACAAACAAAGCAAAACAAACAAACAAACA 1191
Db 1416 TTTGCTCCATATCTAATGACAAACCAATTAACAAACCAATTAACAAACCAATTAACAAACCAATTAAC 1446

Qy 1192 AAAAAACCAACCTTCCAATGTGGGAGAGAGAGTCTTCGAGGCGCGAGTGTGTGCGAC 1251
Db 1447 -----AACCTTCCAATGTGGGAGAGAGAGTCTTCGAGGCGCGAGTGTGTGCGAC 1498

Qy 1252 ACATGCTAGTAGACATCATCTTTAGCAACTCAAGAAACAAAGCAAAACAAACAAACAAAC 1311
Db 1499 ACATGCTAGTAGACATCATCTTTAGCAACTCTTAAAAACAAACAAACAAACAAACAAAG 1558

Qy 1312 AAAATAAGCGC 1323
Db 1559 AAAAAAATGAC 1570

```

RESULT 2

BC033885
 LOCUS BC033885 Homo sapiens, clone IMAGE:5734270, mRNA.
 DEFINITION Homo sapiens, clone IMAGE:5734270, mRNA.
 ACCESSION BC033885
 VERSION BC033885.1 GI:21707343
 KEYWORDS HTC.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 973)
 Strausberg, R.
 Direct Submission
 Submitted (02-JUL-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

REMARK

NTM-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Invitrogen
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 12:40:33 ; Search time 1884.53 Seconds
(without alignments)

11515.871 Million cell updates/sec

Title: US-09-809-545A-1

Perfect score: 1340

Sequence: 1 gggccgcgcctgacacaat.....ggccgaagggttcgctaga 1340

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hct:*

9: gb_est1:*

10: gb_est2:*

11: gb_hct:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estcom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	911.2	68.0	4039	11 BC026312	BC026312 Homo sapi
2	529.2	39.5	973	11 BC033885	BC033885 Homo sapi
3	465.2	34.7	542	9 AI101278	AI101278 EST210567
4	464.2	34.6	587	9 AI885361	AI885361 w192h03.x
5	444.4	33.2	533	12 BF549922	BF549922 UI-R-E0-b
6	417.2	31.1	494	12 BF420112	BF420112 UI-R-BJ2-

C	7	413.4	30.9	479	9	AI291784	AI291784 qm75b08.x
C	8	401.6	30.0	1020	14	BM808707	BM808707 AGENCOURT
C	9	399.4	29.8	457	9	AI193117	AI193117 q669h03.x
C	10	398	29.7	495	9	AI656926	AI656926 tt48a08.x
C	11	391	29.2	774	10	AV729057	AV729057 AV729057
C	12	384.2	28.7	773	10	AV729198	AV729198 AV729198
C	13	383.2	28.6	460	9	AA859573	AA859573 UI-R-E0-b
C	14	379.8	28.3	549	10	AV655025	AV655025 AV655025
C	15	378.2	28.2	451	9	AI193843	AI193843 q67h11.x
C	16	376.8	28.1	428	9	AA460940	AA460940 z261d04.s
C	17	372.8	27.8	449	9	AI133596	AI133596 qt42e01.x
C	18	372.6	27.8	428	10	AV665024	AV665024 AV665024
C	19	368	27.5	464	9	AI332614	AI332614 qp94h10.x
C	20	357.4	26.7	391	9	AI186273	AI186273 q220b08.x
C	21	345.4	25.8	480	9	AI401040	AI401040 th27a12.x
C	22	343.4	25.6	477	10	AW197589	AW197589 xm44g02.x
C	23	338.8	25.3	607	9	AU145000	AU145000 AU145000
C	24	337	25.1	379	9	AI843387	AI843387 UI-M-AQ1-
C	25	334	24.9	593	9	AU118435	AU118435 AU118435
C	26	332	24.8	565	14	W68285	W68285 z835e10.81
C	27	330.2	24.6	449	9	AI374891	AI374891 ta60g12.x
C	28	330.2	24.6	456	9	AI917298	AI917298 ts96g09.x
C	29	328.2	24.5	459	9	AI244212	AI244212 q186b11.x
C	30	328	24.5	430	9	AI850237	AI850237 UI-M-BG1-
C	31	316.6	23.6	395	9	AI717457	AI717457 UI-R-Y0-a
C	32	314	23.4	383	9	AI189407	AI189407 q01h05.x
C	33	313.2	23.4	361	9	AI835011	AI835011 UI-M-AM1-
C	34	311.6	23.3	425	10	BB750925	BB750925 BB750925
C	35	311.2	23.2	723	12	BG306387	BG306387 fm58d11.x
C	36	301.4	22.5	443	9	AA773715	AA773715 af81c05.f
C	37	295.8	22.1	347	9	AI041587	AI041587 ox56g05.x
C	38	295.8	22.1	439	9	AI968176	AI968176 wu14a06.x
C	39	295.8	22.1	439	10	AW589795	AW589795 hg22d04.x
C	40	295.6	22.1	801	9	AU051767	AU051767 AU051767
C	41	295.2	22.0	1589	11	BC002124	BC002124 Mus muscu
C	42	294.8	22.0	439	10	BE501380	BE501380 T44h09.x
C	43	288.2	21.5	521	13	BM183349	BM183349 fv10e03.x
C	44	280	20.9	420	9	AI655094	AI655094 wb57a10.x
C	45	279.2	20.8	538	9	AA818804	AA818804 UI-R-A0-a

ALIGNMENTS

RESULT 1	BC026312	4039 bp	mRNA	linear	HTC 08-APR-2002
LOCUS	BC026312	Homo sapiens, clone IMAGE:4815500, mRNA.			
DEFINITION	BC026312				
ACCESSION	BC026312.1	GI:20070932			
VERSION	BC026312.1				
KEYWORDS	HTC				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens				
REFERENCE	1	(bases 1 to 4039)			
AUTHORS	Strausberg, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-APR-2002) National Institutes of Health, Mammalian				
	Gene Collection (MGC), Cancer Genomics Office, National Cancer				
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,				
	USA				
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov				
COMMENT	Contact: MGC help desk				
	Email: cgapbs@mail.nih.gov				
	Tissue Procurement: Miklos Palkovits, M.D., Ph.D.				
	CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki				
	Toshiyuki and Piero Carninci (RIKEN)				
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)				
	Center, Stanford University School of Medicine, Stanford, CA 94305				
	Web site: http://www.shgc.stanford.edu				
	Contact: (Dickson, Mark) mcd@paxil.stanford.edu				

THIS PAGE BLANK (USPTO)

PN DE20103510-UI.
 XX
 PD 07-JUN-2001.
 XX
 PF 28-FEB-2001; 2001DE-2003510.
 XX
 PR 02-DEC-1999; 99DE-1058160.
 XX
 PA (LION-) LION BIOSCIENCE AG.
 XX
 DR WPI; 2001-368570/39.
 XX
 XX
 PT Gene library containing sequences with specific 3'-ends and no polyA
 PT tail, encoding proteins involved in a wide range of cellular processes
 PT
 XX
 PS Claim 15; Page 174; 251pp; German.
 XX
 CC This invention describes a novel gene library (A) comprises a gene
 CC sequence (or its part) encoding a protein involved in amino acid
 CC synthesis, cellular/energy metabolism, metabolism of
 CC fatty acids/phospholipids, synthesis or breakdown of
 CC purines/pyrimidines/nucleosides/nucleotides, DNA
 CC replication/transcription/translation, or is a transport/binding protein.
 CC (A) are produced that correspond to the 3'-end of mRNA but without the
 CC polyA tail. They can be prepared more efficiently and with less effort
 CC than conventional libraries. AAK53436-AAK54275 represent fragments of the
 CC gene library described in the method of the invention.
 XX
 SQ Sequence 327 BP; 71 A; 65 C; 84 G; 107 T; 0 other;
 Query Match 18.8%; Score 251.8; DB 22; Length 327;
 Best Local Similarity 92.6%; Pred. No. 5.4e-57;
 Matches 299; Conservative 0; Mismatches 17; Indels 7; Gaps 3;
 QY 981 CCACACACTTGTCCA--GCCCCCAGCTAGGCGTTGGTGCATGAATGC-TTTTCGCC 1037
 DB 323 CCACACACTTGTCTCCAAAGCCCCCAGCTAGGCGTTGGTGCATGAATGC-TTTTCGCC 264
 QY 1038 CTTGACCGATGCCAAGCTAGGAGCCATGCTGATGATGTTGGTCTCGTTCTTTTCATT 1097
 DB 263 CTTGACCGATGCCAAGCTAGGAGCCATGCTGATGATGTTGGTCTCGTTCTTTTCATT 204
 QY 1098 GCAGGCTAGTATATACCAAGGGGGATACAAACCGTTTGTCTCCATATTAATGATAAAACC 1157
 DB 203 GCAGGCTAGTATATACCAAGGGGGATACAAACCGTTTGTCTCCATATTAATGATAAAACC 144
 QY 1158 ATTAACAACAAGCAAAACAAACA-----AAAAACAAACCAACCTTCCAATGTG 1213
 DB 143 ATTAACAACAAGCAAAACAAACA-----AAAAACAAACCAACCTTCCAATGTG 84
 QY 1214 GGGAGAGAGAGCTTTTCCGAGCGCGAGTGTTCGACATGCAGTAGGACATCATT 1273
 DB 83 GGGAGAGAGAGCTTTTCCGAGCGCGAGTGTTCGACATGCAGTAGGACATCATT 24
 QY 1274 AGCAACTCAAGAAACAAACGAAA 1296
 DB 23 AGCAACTCAAGAAACAAACGAAA 1

Search completed: March 15, 2003, 12:51:47
 Job time : 329.202 secs

PT human RNA binding proteins which play a role in cancer, immune disorders and developmental disorders -

XX Examples: Columns 51-52; 39pp; English.

CC The present invention provides human RNA binding proteins (RNABP) and polynucleotides encoding the proteins. The polypeptides are useful for modulating the expression of human RNA binding proteins (RNABP) which play a role in cancer, immune disorders and developmental disorders. Disorders associated with a decrease of RNABP include: cancers such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; immune disorders such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxicity, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hyperosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjogren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, trauma, X-linked agammaglobulinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated immunoglobulin (Ig)-A deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, and immunodeficiency associated with Cushing's disease; and developmental disorders such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucopolysaccharidosis, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, sensorineural hearing loss, and any disorder associated with cell growth and differentiation, embryogenesis, and morphogenesis involving any tissue, organ, or system of a subject, e.g., the brain, adrenal gland, kidney, skeletal or reproductive system. The present sequence represents a cDNA encoding the RNABP-2 polypeptide.

XX Sequence 1506 BP; 452 A; 362 C; 377 G; 315 T; 0 other;

Query Match 20.6%; Score 275.6; DB 21; Length 1506;
Best Local Similarity 60.2%; Pred No. 5e-63;
Matches 601; Conservative 0; Mismatches 354; Indels 43; Gaps 7;

QY 222 GACAGATGATCGCGCCGACGACGAGCGGCGAGC---CCCAGACACACCTTCTGAAAAACAC 278
DB 452 GACAGAGGTGGAGCAGACAGACGAGCGGCGAGCAGTACACAGACACAAAGTAGTGAATAATTC 511
QY 279 AGAAACAAGTCCAGCCCAAGCGGTGATGTCTCAACATCCCTTCGGGTCGCGGA 338
DB 512 AGAGAGTAAATCTACCCGAAACGGCTGATGTCTCTAATATTCCTTCCTTCGGGTCGCGGA 571
QY 339 TCCAGACCTCCGACAAATGTTGGCCAAATTTGGTAAATATTTAGATGTTGAAATATTTT 398
DB 572 CCTTGACCTCCGCGAGATGTTGGGCGAGTTGGCAAAATCTAGATGTAGAATAATCTT 631
QY 399 TAATGAGCGGGCTCGAAGGGATTTGGTTTCGTAACCTTTGAAAAATAGTGGCGATCGGA 458

Db 632 TAATGAACCTGGCTTAAGGGATTTCGGTTTCGTAATCTTCGAGATAGTCTGATGCAGA 691
QY 459 CAGGCGCAGGAGAGAAATTCACCGGTACCGTGTAGAGGCCGTAAATCGAGTTAATAA 518
DB 692 CAGGCGCAGGAGAGAAATTCACCGGTACCGTGTAGAGGCCGTAAATCGAGTTAATAA 751
QY 519 TGGCAGCAGCAGCGGTGATGACTTAATAAAGGCCGTGAACCCCTACACCAATGGCTGGAA 578
DB 752 TGTACAGCAGCGTGTATGACCAATGAAGAGATGGTGCACACCATATGCAAAATGTTGGAA 811
QY 579 ATTAATCCAGTTTGGCGCGGTCTACAGCCCGGACTTCTATGCAG-----GCA 628
DB 812 ATTAAGCCCAAGTATGAGAGCTGTATATGTCGGAGTTATATGACAGATCCAGCTTCA 871
QY 629 CGGTGCTGTTTGTCCAGGCCAACACAGAGGGATTCCTTCATGTA--CAGTGGCCCCAGTTC 686
DB 872 AGCAGATGTGTCCTAGGCAATGATCAGCAGTGCCTTATCAGGAAGAGGGGTATCAA 931
QY 687 ACTTGTATATCTTCTGCAATGCTGCTTTCATA---TCGCGCCGCACTCTGTCAGC 743
DB 932 CACTTACATTCCTTAAATCAATTCCTGCTTCCCTTACCCTACTGACGCCACACGGGAGC 991
QY 744 TGCATACCGAGGGCTCACCTTCGAGGCCGTGTCGACCGTGTACAAACCTTTCAGAGC 803
DB 992 CGCTTTCAGAGGAGCCCATTTGAGGGCAGAGGGCGGACAGTATATGTCAGTCCGA-- 1049
QY 804 TGGCGGCCCCACCCCAATCCCGGCTATGCGGAGTAGTGTATCAAGAGCCAGTGTGA 863
DB 1050 -GCGGTACTCTCAACAGCCATCCCGCTATCAGGGGTGATATGAGCTACAG---- 1104
QY 864 TGGCAATAAATGCTACAGGGTGTAGCTGTACCTGTCAGCCGACGCCACCCCTGTC 923
DB 1105 -----ATATGCACAGCTGCTACTGCAACCGCAGCCACCGTGTCTGCAGC 1149
QY 924 CACTGTGCTGCTACAGTACAGTACGAGTACGAGTATGCTGCTGCCGACCCCTACACCA 983
DB 1150 CGTTCAGAGCCCTTACAGTACGCTTATGCGAGGTGTACAGCCGACCCCTACCATGC 1209
QY 984 CACACTTGTCTCAGCCCCCACCCTACGCGTGTGGTGCATGAATGCTTTTGGGCCCTTGC 1043
DB 1210 C---CTTGGCCCTGCTAGTATGAGTGTGGCTGTGGCGAGTTTATATACCGAGTGG 1266
QY 1044 CGATGCCAAGCTAGGAGCCATGCTGATGATGTGGTCTGTTCTTCTTCTTCTTCTTCT 1103
DB 1267 CTACAGCCGATTTTGGCCCTTACTGAGTGCAGTGCAGCCCTGCAAAATGGGACAGCCCC 1326
QY 1104 TAGTATATACCAAGGGGATACACCCGTTTGTCTCATATTTAAATGATAAAACCATTA 1163
DB 1327 CAGTTTCATGAGCCCTGCTTGTGCAATATTTACTAGTAGAGGAACTCTATAGCAAGATGA 1386
QY 1164 CAAACAAGCAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 1201
DB 1387 AGAGGAAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 1424

RESULT 15

AAK54001/C

ID AAK54001 standard; cDNA; 327 BP.

XX AAK54001;

XX 16-NOV-2001 (first entry)

XX Murine transcription associated protein encoding cDNA SEQ ID 566.

DE Murine; liver; gene library; amino acid synthesis; binding protein;

XX cell metabolism; energy metabolism; fatty acid metabolism; synthesis;

KW phospholipid metabolism; purine; pyrimidine; nucleoside; nucleotide;

KW replication; transcription; translation; transport protein; ss.

XX Mus musculus.

OS

XX

26-DEC-2000; 2000WO-US34263.
21-JAN-2000; 2000US-0488725.
25-APR-2000; 2000US-0552317.
09-JUL-2000; 2000US-0598042.
19-JUL-2000; 2000US-0620312.
03-AUG-2000; 2000US-0653450.
14-SEP-2000; 2000US-0662191.
19-OCT-2000; 2000US-0693036.
29-NOV-2000; 2000US-0727344.
(HYSE-) HYSEQ INC.
Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
Zhao QA, Zhou P, Goodrich R, Drmanac RT;
WPI: 2001-442253/47.
P-PSDB; AAM40092.
Novel nucleic acids and polypeptides, useful for treating disorders
such as central nervous system injuries -
Claim 1; SEQ ID NO 1451; 10078pp; English.
The invention relates to human nucleic acids (AAI57798-AAI61369) and
the encoded polypeptides (AAM38642-AAM42213) with nootropic,
immunosuppressant and cytostatic activity. The polynucleotides are useful
in gene therapy. A composition containing a polypeptide or polynucleotide
of the invention may be used to treat diseases of the peripheral nervous
system, such as peripheral nervous injuries, peripheral neuropathy and
localised neuropathies and central nervous system diseases, such as
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
utilisation of the activities such as: Immune system suppression,
Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
and thrombolytic activity, cancer diagnosis and therapy, drug screening,
assays for receptor activity, arthritis and inflammation, leukaemias and
C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
specification.
XX
SQ Sequence 1011 BP; 260 A; 314 C; 267 G; 170 T; 0 other;
Query Match 21.4%; Score 286.6; DB 22; Length 1011;
Best Local Similarity 64.6%; Pred. No. 5e-66;
Matches 574; Conservative 0; Mismatches 219; Indels 96; Gaps 6;
QY 155 CCTACACAGACGCTCGGAGCAGAGCGTATTATACAGCGCACAGCGTCTCCGGCACCG 214
DB 157 CCCACCCCGACGACGCGTCCGAGGCCAGCACACAGCGCCATCGCGGGACCCAGACAG 216
QY 215 CCACACAGACAGATGATCGCCGCCACGACGCGCCAGCCACACACCTTCTGAA 274
DB 217 TGGCCGACAGACAGAGCGCGACAGCGACGACGCGCGCT---CCACCCCTCCGACC 273
QY 275 ACACAGAAACAGTCCAGCCCAAGCGGCTGCATGTGTCCAAACATCCCTTCGGGTCC 334
DB 274 CTACAGAGAAGCAG---CAGCCCAAGCGGCTACACGTCTCCACATCCCTTCGGGTCA 330
QY 335 GGGATCCAGACCTCCGACAAATGTTTGGCCAATTTGGTAAATATATAGATGTGAAATTA 394
DB 331 GGGACCCCGACTTGGCGCAATGTTCCGGCAATTCGGAATAATTTAGACGTGGAGATCA 390
QY 395 TTTTAAATGAGCGGGCTCGAAGGATTTGGTTTTCGTAACCTTCGAAATAGTGGGATG 454
DB 391 TTTTAAACGAGCGGGCTCCAAGGGTTTGGGTTTGTAACTTTTGAACACTAGCTCAGATG 450
QY 455 CGGACAGGGCGAGGAGAAATTTGACCGTACCGTGTAGAGGGCGGTAAATTCGAGGTTA 514
DB 451 CTGACCGAGCCCGGAGAGAGCTGAATGGGACGATCGTAGGGGACGGAATAATTGAGGTCA 510
QY 515 ATAATCGACAGACGCGGTGATGACTAATAAAGAGCCCGTGAACCCCTACACCAATGGCT 574

DB 511 ATAATGCCAGCGCCCGAGTGATGACCAAGAAGACGGGAACCCCTACCAACGGCT 570
QY 575 GGAATTAATCAGTTGTGGCGGCTCTACAGCCCCGACTTCTATGAGGACACGGTGC 634
DB 571 GGAAGCTAAATCAGTGGTCGGCGCAGTCTACGGGCTGAATTTCTATGCG----- 621
QY 635 TGTGTGCCAGGCAACACAGGAGGATCTTCCATGTACAGTGGCCCCCAGTTCACTTGTAT 694
DB 622 ----- 621
QY 695 ATACTTCTGCAATGCTGCTTTCCATATCCGGCCGCCACTGCTGCAGTGCATACCGAG 754
DB 622 -----TGACGGGGTTCCCTTACCCACACCGGCACAGCGTTGCTTACCGGG 669
QY 755 GGGCTCACCTTCGAGGCGGTGGTCGCACCGTGTACAAACCTTCAGAGCTGCGGCGCCCC 814
DB 670 GCGCACATCTTCGGGGCCGGGGCCGGTGTATATATACATTTTCGGGGTGGCGCCACCCC 739
QY 815 CACCCCAATCCCGGCTATGCGGAGTAGTGTATCAAGAGCCAGTGTATGGCAATAAT 874
DB 730 CACCCCAATCCCGGCTATGCGGAGCGGTGTATCAGATGGATTTTATGTGCTGA-- 787
QY 875 TGCTACAGGCTGTACGCTGCTACCGCTACCGCCAGCCCTGCCACTGCTGCTG 934
DB 788 -GATTTATGGAGCTACGCAGCTACAGATACGTCAGC---CCGCTGCAGCGGGCGCAG 843
QY 935 CCTACAGTACAGATTACGACAGATTTA---TGCTGCCGACCCCTACCAACACACTTGG 991
DB 844 CTTACAGCGACAGTTCAGCGCAGAGTCTACGCGCTGCCGACCCGTACCATCACCATCG 903
QY 992 CTCCAGCCCCCACCTACGCGGTTGGTGCCATCAATGCTTTTCGGCCCTT 1040
DB 904 GCGCCGCGGCGACCTACAGCATTTGGAACCATGTGAAACCTTCCACCGTT 952
RESULT 14
AA246827
ID AA246827 standard; cDNA; 1506 BP.
XX
AC AA246827;
XX
DT 11-APR-2000 (first entry)
XX
DE Human RNA binding protein (RNABP)-2 encoding cDNA (clone 1250374).
XX
KW RNA binding protein; RNABP; cancer; immune disorder; AIDS; human;
KW developmental disorder; acquired immunodeficiency syndrome; RNABP-2;
KW inflammation; allergy; diabetes mellitus; seizure disorder; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 201..1301
FT /*tag= a
FT /product= "RNABP-2"
XX
PN US6020164-A.
XX
PD 01-FEB-2000.
XX
PF 21-OCT-1998; 98US-0176657.
XX
PR 21-OCT-1998; 98US-0176657.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Bandman O, Corley NC, Lu DAM, Baughn MR, Tang YT, Guegler KJ;
XX
DR WPI: 2000-146885/13.
DR P-PSDB; RAY56850.
XX
PT Isolated and purified polynucleotide for modulating the expression of

XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-052317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Raundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI; 2001-442253/47.
 DR P-PSDB; NAM41879.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1; SEQ ID NO 5024; 10078pp; English.
 XX
 SS The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (NAM38642-NAM42213) with nontropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 939 BP; 148 A; 248 C; 297 G; 246 T; 0 other;
 XX

Query Match 21.4%; Score 286.6; DB 22; Length 939;
 Best Local Similarity 64.6%; Pred. No. 4.8e-66;
 Matches 574; Conservative 0; Mismatches 219; Indels 96; Gaps 6;
 6;

QY 155 CCTACACAGCAGCTCGGAGCAGCGCTTATACAAAGCGCAGCGCTCTCGGCGACG 214
 DB 875 CCACCCCGAGCGCAGGCTCGAGGCCAGCAGCAGCGCCATCGCGGGAGCCAGACAG 816
 QY 215 CCACACAGCAGATGATGCGGCCCGCCAGCGCGCCAGCCAGCACAACCTTCTGAAA 274
 DB 815 TGCGCGCAGACAGCAGCGCGGCAGCAGCAGCAGCGCGCT---CCACCCCTCGGACC 759
 QY 275 ACACAGAAACAGTCCGAGCCCGAGCGGCTGATGTGTCACATCCCTTCCGTTCC 334
 DB 758 CTACAGAGAAGCAG---CAGCCCAAGCGGCTACACGCTCTCCACATCCCTTCCGTTCA 702

QY 335 GGATCCAGACCTCCGACAAATGTTTGGCCAATTTGGTAAATATATAGATGTGAAATTA 394
 DB 701 GGGACCCCGCATTTGCGGCAAAATGTTTGGGCAATTCGGAATAATTTTAGACGTGGAGATCA 642
 QY 395 TTTTAAATGAGCGGGGCTCGAAGGGAATTTGGTTTCTGTAACCTTTGGAATAATAGTGGCGATG 454
 DB 641 TTTTAAACGAGCGGGGCTCCAGGGGTTTGGGTTTGTAACTTTTGAACCTAGCTCAGATG 582
 QY 455 CGGACAGGGCGAGGAGAAATTTGCACGGTACCGTGTAGAGGCCGTAAATCGAGGTTA 514
 DB 581 CTGACCGAGCCCGGAGAGAGCTGAATGGGACGATCGTAGAGGAGCGAATAATGAGGTCA 522
 QY 515 ATAATGGACAGCAGCGGTGATGACTAATAAAGAGCGGTGAACCCCTACACCAATGGCT 574
 DB 521 ATAATGCCACGCGCGGCTGATGACCAAGAAGACGGGGAACCCCTACACCAACGGCT 462
 QY 575 GGAATAATAATCCAGTTGTGGGCGGTCTACAGCCCCGACTTCTATGACAGGCACGGTGC 634
 DB 461 GGAAGCTAATAATCGTGTGGGCGAGTCTACGGGCTGNAATCTATGCAAG----- 411
 QY 635 TGTGTGCCAGGCCAACACAGGAGGATCTTCCATGTACAGTGGCCCCAGTTCACTTGTAT 694
 DB 410 ----- 411
 QY 695 ATACTTCTGCAATGCTGGCTTTCCATATCCGGCGCCACTGCTGCAGCTGCATACCGAG 754
 DB 410 -----TGACGGGGTTCCCTTACCCACACCGGCACAGCGGTTCCTTACCGGG 363
 QY 755 GGGCTACCTTTCCAGGCGGTGTCGACACCGTGTACAAACACTTTCAGAGCTGCGGCGCCCC 814
 DB 362 GGCACATCTTCGGGCGGGGCGGCGCGGTGTATATATACATTTTCGGGCTGCGCCACCCC 303
 QY 815 CACCCCAATCCCGGCTATGGCGAGTAGTGTATCAAGAGCCAGTGTATGGCAATAAAT 874
 DB 302 CACCCCAATCCCGGCTATGGCGAGCGGTGTATCAAGAGTGTATTTATGGTGTGA-- 245
 QY 875 TGCTACAGGTTGTACGCTGCATACCGCTACCGCAGCCACCCCTGCACCTGCTGCTG 934
 DB 244 -GATTTATGGAGGCTACGACGCTTACAGATACGCTCAGC---CGCTGACAGCGCGGCGAG 189
 QY 935 CCTACAGTGACAGTTACGGAGCGAGTTTA---TGCTGCCACCCCTACACACACACACTTG 991
 DB 188 CCTACAGGACAGTTACGGAGAGTCTACGAGCTGCGGACCCGCTACCATCACACCATCG 129
 QY 992 CTCAGGCCCCCAGCTACGCGGTTGGTCCCATGAATGCTTTTGGGCCCTT 1040
 DB 128 GGCCCGCGGCGACCTACAGCATTGGAACCATGTGAACCTTCCACCCGTT 80

RESULT 13
 AAI59248
 ID AAI59248 standard; cDNA; 1011 BP.
 XX
 AC AAI59248;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 XX Human polynucleotide SEQ ID NO 1451.
 DE
 XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200153312-A1.
 XX
 PD 26-JUL-2001.
 XX

Qy 1235 GGCCCGAGTGTTCGACACATGCAGTAGGACATCACTT 1272
|||||
Dd 38 GGCCTGACTATTGGAATACATGCAGTAGTGATCATCACTT 1

RESULT 11	
AAI61034/c	
ID	AAI61034 standard; cDNA; 939 BP.
XX	
AC	AAI61034;
XX	
DT	22-OCT-2001 (first entry)
XX	
DE	Human polynucleotide SEQ ID NO 5023.

Best Local Similarity		64.6%	Pred. No. 4.8e-66;
Matches 574; Conservative		0; Mismatches 219; Indels	96; Gaps
QY	155	CCTACACAGACGCACTCGGAGCAGACGCGCTTATACAAAGCCACAGACGCGTCTCCGGCACCG	214
Db	875	CCACCCCGACGACCGAGGCTCGAGGCGACGACACAGCCATCGCGGGACCCACGACAG	816
QY	215	CCACACAGACAGATGATGCGCGCCCGACCGACGCGCCAGCCCCACAGACACAACTTCTGAAA	274
Db	815	TGCGCGACAGACGAGCGCGCACAGACGCGACGACGACGCGCT--CCACCCCTCCGACC	759
QY	275	ACACAGAAAAAAGTCCGAGCCCAAGCGCGGTGCATGTGTCAACATCCCTTCGGGTTC	334
Db	758	CTACAGAAAGCAG--CAGCCCAAGCGGTACACGCTCTCCAACATCCCTTCGGTTCA	702
QY	335	GGGATCCAGACCTCCGACAAATGTTTCGCGCAATTTGCTAAATATTAGATGTTGAAATTA	394
Db	701	GGGACCCGCACTTCGGCAAAATGTTTCGGCAATTCGGAATAATTTAGACGTGGAGATCA	642
QY	395	TTTTTAATGAGCGGGGCTCGAAGGGATTTGGTTTCGTAACTTTCGAAAATAGTGCAGATG	454
Db	641	TTTTTAACGAGCGGGGCTCCAAGGGTTTTCGGTTTGTAACTTTGAACTAGCTCAGATG	582
QY	455	CGGACAGCGCGAGGGAGAAATTTGCACGGTACCGTGTAGAGGCCGTAAATCGAGGTTA	514
Db	581	CTGACCGAGCCGCGGAGGAAAGCTGAATGGGACGATCGTGTAGAGGACGGAATTTGAGTCA	522
QY	515	ATAATGGCAGACGCGTGTATGACTAATAAAGGCGGTGAACCCCTACACCAATGGCT	574
Db	521	ATAATGCCACGGCCCGAGTGATGACCAACGAAGAGACGGGAACCCCTACACCAACGGCT	462
QY	575	GGAAATTAATTCACAGTTGTGGGGCGGCTCTACAGCCCGCACTTCTATGAGAGCACGGTGC	634
Db	461	GGAGCTAAATCCAGTGTGCGGCGAGTCTACGGGCTGNAATCTATGAG-----411	
QY	635	TGTTGTGCGAGGCCAACCAAGAGGGAATCTTCCATGTACAGTGAGGCCCGAGTTTCACTTGTAT	694
Db	410	-----TGAGGGGGTTCCTTACCCACACCGGCACAGCGGTGCTTACCGGG-----411	
QY	695	ATACTTCTGCNATCCCTGGGCTTTCCATATCCGGCGCGCACTGCTGCAGCTGCATACCGAG	754
Db	410	-----TGAGGGGGTTCCTTACCCACACCGGCACAGCGGTGCTTACCGGG363	
QY	755	GGGCTCACTTCGAGGCGGTGTCGCACCGTGTACAAACCTTTCAGAGCTGCGGCGCCCC	814
Db	362	CGGCACATCTTCGGGCGGGGCGGGCGGTGTATAATACATTCGGGCTGGCCACCCCT	303
QY	815	CACCCCAATCCCGGCTATGGCGGAGTATGATATCAAGAGCCAGTGTATGGCAATAAT	874
Db	302	CACCCCAATCCCGACTTACGGAGCGGTGCTGTATCAGGATGATTTATGGTGCTGA--245	
QY	875	TGCTACAGGTGGTTACGCTGCATACGGCTACGCCAGCCGCACTCCGCACTGCTGCTGT	934
Db	244	-GATTTATGGAGGCTACGGACGCTACAGATACGCTCAGC---CCGCTGAGGGGGCGGAG	189
QY	935	CCCTACGTGACAGTTACGGACGAGTTTA---TGCTGCCGACCCCTACACACACACACTTG	991
Db	188	CCCTACGCAAGTTACGGCAGAGTCTACGAGCTGCCGACCCGTCACCATCACCATCG	129
QY	992	CTCCAGCCCCACCTACGGCGTTTGGTGCCATGAATGCTTTTTCGGCCCTT1040	
Db	128	GGCCCGGGGACCTACAGCATTTGGAACCATGTGAAACCTTCCACCGTT80	
RESULT 12			
AAI61035/c			
ID AAI61035 standard; cDNA; 939 BP.			
XX	AAI61035;		
AC			
DT	22-OCT-2001 (first entry)		
XX	Human polynucleotide SEO ID NO 5024.		

Human polynucleotide SEO ID NO 5024.


```
Qy 739 GCAGCTGCATACCGAGGGCTCACTTCGAGCGCGTGTGCGACCGTGTACAAACACCTTC 798
Db 1 GCGGCGCGCTACCGAGGGCGCACTGCGAGCGCGGTGCGACCGTGTACAAACACCTTC 60
Qy 799 AGAGCTGGGCGCGCCACACCCCAATCCGCGCTATGCGGAGTAGTATCAAGAGCCA 858
Db 61 AGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Qy 859 GTGTATGCAATAAATTGCTACAGGCGTGTACGCTGCATACCGCTAGCGCCAGCCACCC 918
Db 121 TTTTATGG--TGCAGACATTTATGTTGTTATGCTGCATACCGCTAGCGCCAGCGCTACC 177
Qy 919 CTTGCACTGCTGCTGCTCAAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 976
Db 178 CTTGCACTGCTGCTGCTCAAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 237
Qy 977 ACCACCAACACTGCTCCAGCGCCCACTACGCGGTTGGTCCATGAATGCTTTTGGCG 1036
Db 238 GTCACCACTCAGTCTCCAGCGCCCACTACGCGGTTGGTCCATGAATGCTTTTGGCG 297
Qy 1037 CTTGACCGATGCCAAGACTAGGAGCCATGCTGATGATGTTGGTCTCGTTCTTCTTCAT 1096
Db 298 CTTGACCGATGCCAAGACTAGGAGCCATGCTGATGATGTTGGTCTCGTTCTTCTTCAT 357
Qy 1097 TGCAGGCTAGTATATACCAAGGGGATACAAACCGTTTGTCTCCATATTAATGATAAACC 1156
Db 358 TGCAGGCTAGTATATACCAAGGGGATACAAACCGTTTGTCTCCATATTAATGATAAACC 417
Qy 1157 CATTAACAAACAGCAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 1216
Db 418 CATAAA-----AACCTTCCAATGTGGG 440
Qy 1217 AGAGAGGAAGCTTTCGAGGCGCGAGTGTGCGACACATGC 1257
Db 441 AGAAGGAAGCTTTCGAGGCGCGAGTGTGCGACACATGC 481
RESULT 10
ABK34530/C
ID ABK34530 standard; cDNA; 539 Bp.
XX
AC ABK34530;
XX
DT 08-MAY-2002 (first entry)
XX
DE Human cDNA for novel secreted protein, SEQ ID 299.
XX
KW Human; ss; gene; secreted protein; immune deficiency; viral infection;
KW bacterial infection; fungal infection; autoimmune disorder; burn;
KW rheumatoid arthritis; multiple sclerosis; autoimmune thyroiditis;
KW diabetes; allergy; asthma; neurodegenerative disease; Crohn's disease;
KW Alzheimer's disease; Parkinson's disease; liver fibrosis; tumour;
KW coagulation disorder; haemophilia; inflammatory disorder; ulcer;
KW tissue regeneration; wound healing; haematopoiesis; myeloid deficiency;
KW lymphoid cell deficiency.
XX
OS Homo sapiens.
XX
PN WO20017290-A2.
XX
PD 18-OCT-2001.
XX
PF 29-MAR-2001; 2001WO-US10295.
XX
PR 06-APR-2000; 2000US-194941P.
XX
PA (GEMY ) GENETICS INST INC.
XX
PI Wong GG, Clark HF, Fichtel K, Agostino MJ, Howes SH, Resnick RJ;
PI Gulukota K, Graham JR;
XX
XX WPI; 2002-179323/23.
XX
```

```
PT Six hundred and twenty five polynucleotides derived from a variety of
PT human tissue sources which encode secreted proteins, useful for
PT treating immune deficiencies and disorders such as autoimmune disorders
XX
XX Claim 1; Page 153; 339pp; English.
XX
CC The invention relates to 625 polynucleotides which have been derived from
CC a variety of human tissue sources and which encode novel secreted
CC proteins, their complements and sequences that hybridise to them.
CC Also included are a vector comprising the polynucleotide, a host cell
CC transformed with the vector, the proteins encoded by the
CC polynucleotides, antibodies that bind to the proteins and identification
CC of modulators of the proteins or the expression of the polynucleotide.
CC The polynucleotides can be used as probes for the identification
CC and isolation of full length cDNA and genomic DNA. The polynucleotides
CC and proteins can also be used as nutritional supplements. The protein
CC is useful in the treatment of various immune deficiencies and disorders
CC such as viral infections, bacterial infections, fungal infections,
CC autoimmune disorders (e.g. rheumatoid arthritis, multiple sclerosis,
CC autoimmune thyroiditis and diabetes) and allergic reactions and
CC conditions (e.g. asthma). They are also useful for treating
CC neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
CC disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
CC useful for tissue regeneration, for wound healing and in the treatment
CC of burns, incisions and ulcers. The proteins are also useful for
CC regulating haematopoiesis, for treating myeloid or lymphoid cell
CC deficiencies. The present sequence is one of the 625 cDNA sequences
CC encoding a secreted protein.
XX
SQ Sequence 539 BP; 132 A; 132 C; 156 G; 119 T; 0 other;
Query Match 21.6%; Score 290; DB 24; Length 539;
Best Local Similarity 76.1%; Pred. No. 4.6e-67;
Matches 440; Conservative 0; Mismatches 45; Indels 93; Gaps 3;
```

Qy 748 TACGAGGGGCTACCTTCGAGGCGGTGTCGACCGTGTACAAACACCTTCAGAGCTGCG 807
Db 538 TACGAGGGGCGCACCTTCGAGGCGCGGTGTCGACCGTGTACAAACACCTTCAGAGCTGCG 479

Qy 808 GCGCCCCCACCACCAATCCGCGCTATGGCGGAGTAGTGTATCAAGAGCCAGTGTATGGC 867
Db 478 GCGCCCCCAGCGAGTCCGCGCTACGCGGCTGTGTTTACCAGGATGGATTTATGG- 420

Qy 868 AATAAATTGCTACAGGCTGTTACGCTGCATACCGCTACGCGCCAGCCACCCCTGCCACT 927
Db 419 --TGCAGACATTTATGTTGTTATGCTGCATACGCTACGCGCCAGCTACCCCTGCCACT 362

Qy 928 GCTGCTGCTACAGTGA----- 944
Db 361 GCCGCTGCCTACAGTGCAGAGAAATCAGTTCTGTTGTCAGCAGATGAAATTTCTTGT 302

Qy 945 -----CAGTTACGAGAGTGTATGCTGCGCGAGCCCTACCCACACACCTTGCTC 994
Db 301 AACACCTCTGCGATGATGCGGCTGTTGTTGCTGCGAGCTGTTTGCACCTTTGACTGATGCAAGA 182

Qy 995 CAGCCCCCACCCTACGCGGTTGGTGCCTATGAATGCTTTTGGCGCCCTTGACCGATGCCAAGA 1054
Db 241 CAGCCCCCACCCTACGCGGTTGGTGCCTATGAATGCTTTTGCACCTTTGACTGATGCAAGA 182

Qy 1055 CTAGAGCCATGCTGATGATGTTGGTCTCGTTCTTCTTCAATTCGAGCTAGTATATACC 1114
Db 181 CTAGAGCCATGCTGATGATGTTGGTCTCGTTCTTCTTCAATTCGAGCTAGTATATACC 122

Qy 1115 AAGGGGATACACCGTTTCTGCTCCATATTAATGATAAACCATTAAACAAACAGCAA 1174
Db 121 GAGGGGGATACACCGTTTCTGCTCCATATTAATGATAAACCATTAAACAAACAGCAA 74

Qy 1175 AAAACAAAACAAAACAAAACAAAACAAACCTTCCCAATGTGGGGAGAGAGAGCTTCCGA 1234
Db 73 -----NACCTTCCAATGTGGGGAGAGAGAGCTTCCGA 39

XX WPI; 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX Claim 1; SEQ ID 513; 2537pp + CD ROM; English.
XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX Sequence 893 BP; 225 A; 228 C; 236 G; 199 T; 5 other;
SQ

Query Match 24.9%; Score 334; DB 22; Length 893;
Best Local Similarity 85.1%; Pred. No. 1.1e-78;
Matches 479; Conservative 0; Mismatches 71; Indels 13; Gaps 9;
QY 1 GCGGCGCCCTGACACAATGGCTCAG-CTTATGCTCAGCGAGTTGCTCCACCCAG 59
DB 263 GCGCTGCCCTGACACAATGGCTCAGCTTACGCTTCGGCCAGTTTGTCCCCGCG 322
QY 60 AATGGCATCTT-GCAGATAACAGCGCCCTCATCCCATCCGCGCCAGAG-ACACCGCG 117
DB 323 AACGGTATCCCGGGAATACAGCGCCCTCATCCCAACCCGCGCCAGATACAGGC 382
QY 118 CAGCCACTGTCCCG-CCACACATTAACCTTG-ATCCTCTTACAGACGCACTGGAG 175
DB 383 CAGACCACGGTTCGCGAGCACACATTAACCTGTATCCCTCCGCGCCAGCGCACTCGAG 442
QY 176 CAGAGC--GCTTATACAAGCGACAG-CGGTCTCGGCGACCGCCACACAGACAGATGAT 231
DB 443 CAGAGCCGCGGACAGCAGCGCTCAGACGCTCTCTGGACCGCCACAGGAGATGAC 502
QY 232 GCGGCCCGGACCGACCGCCAGCCGCGCCAGACACAACTTCTGAAAACACAGAAAACAGTCC 291
DB 503 GCAGCAGCAGGATGCGCAGCGCCGCGCCAGACACACACTTCTGAAAACAGGAAACAGTCT 562
QY 292 CAGCCCAAGCGGTGATGTGTCAACATCCCTTCCGTTCCGGGATCCAGACCTCCGA 351
DB 563 CAGCCCAAGCGGTGATGTGTCAACATCCCTTCCGTTCCGGGATCCGGACCTCAGA 622
QY 352 CAATGTTTCGCAATTTGCTTAAATATTAGATCTTGAAATTTTAAATAGCGGGC 411
DB 623 CAATGTTTCGCAATTTGCTTAAATATTAGATCTTGAAATTTTAAATAGCGGGC 682
QY 412 TCGAAGGATTTGGTTTCGTAACCTTTTCGAAAATAGTCGGATGC--GGACAGGCGCCAGGG 469
DB 683 TCAAGGATTTGGTTTCGTAACCTTTTCGAAAATAGTCGGATGCAGGAGG 742
QY 470 AGAAATTGCAAGTACCGTGGTGA--GGCGCGTAAATCAGGTTTAATATGCGACAGC 527

DB 743 AGAAATTACAGGACCGGTGTTAAANGGCCCGTAAATCGAGGTAAATAATGCCCCC 802
QY 528 ACGGTGATGACTAATAAAAAGG 550
DB 803 ACGTGTAAATGACCAATTAAGAAG 825
RESULT 9
AAS93633
ID AAS93633 standard; cDNA; 481 BP.
AC AAS93633;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #29437.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
PD 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR P-PSDB; ABG29446.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 1; SEQ ID No 29437; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 481 BP; 100 A; 145 C; 123 G; 113 T; 0 other;
Query Match 23.8%; Score 319; DB 23; Length 481;
Best Local Similarity 81.4%; Pred. No. 8.4e-75;
Matches 424; Conservative 0; Mismatches 55; Indels 42; Gaps 3;


```

Db      842 CAGACCAAGGTTCCCGAGACACATTAACCTGACCTCCCGCCCAAGCAGCAGCTCCGAG 783
Oy      176 CAGAGC---GCTTATACAGGCGACAG--CCGTCTCCGACCCGACACAGACAGATGAT 231
Db      782 CAGAGCCCGGGGACAGAGCGCTCAGACCGTCTCTGCGACCGCACACAGACAGATAC 723
Oy      232 GCCGCCCCGACCGACGCGCCAGCCCCAGACACACCTTCTGAAAAACAGAAAAACAAGTCC 291
Db      722 GCACGACCGAGCGATGGCGACCGCCAGACACACCTTCTGAAAAACAGAAAAACAAGTCT 663
Oy      292 CAGCCCAAGCGCTGACATGTGTCCAAACATCCCTTCGCGTCCGCGGATCCAGACCTCCGA 351
Db      662 CAGCCCAAGCGCTGACATGTGTCCAAACATCCCTTCGCGGATCCGCGGATCCAGACCTCA 603
Oy      352 CAAATGTTGGCCAAATTTGTAATAATTAAGATGTTAAATTTTAAATGAGCGGCGC 411
Db      602 CAAATGTTGGTCAATTTGTAATAATTTAAATGTTAAATTTTAAATGAGCGGCGC 543
Oy      412 TCGAAGGATTTGGTTTCTGTAACCTTCGAAAAATAGTGGGATGGCGACAGGGCGAGAG 471
Db      542 TCGAAGGATTTGGTTTCTGTAACCTTCGAAAAATAGTGGGATGGCGACAGGGCGAGAG 483
Oy      472 AAATGCAAGGTACCGGTGTAGAGGCGCGTAAATCGAGTTAAATGCGACAGCAGC 531
Db      482 AAATGCAAGGTACCGGTGTAGAGGCGCGTAAATCGAGTTAAATGCGACAGCAGC 423
Oy      532 GTGATGACTAATAAAAGGCGGTGAACCCCTACACCAATGCG 573
Db      422 GTATGACAAATTAATAAGACCGCGGCTCTGCTGGAGTGC 381

```

RESULT 6

AAKS3229/c
ID AAK53229 standard; cDNA; 1164 BP.

AC AAK53229;

DT 06-NOV-2001 (first entry)

DE Human polynucleotide SEQ ID NO 2758.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KM tissue growth factor; immunomodulatory; cancer; leukemia;

KM nervous system disorder; arthritis; inflammation; ss.

XX Homo sapiens.

OS Homo sapiens.

PN MO200157190-A2.

XX 09-AUG-2001.

PD 05-FEB-2001; 2001MO-US04098.

XX 03-FEB-2000; 2000US-0496914.

PR 27-APR-2000; 2000US-0560875.

PR 20-JUN-2000; 2000US-0598075.

PR 19-JUL-2000; 2000US-0620325.

PR 01-SEP-2000; 2000US-0654936.

PR 15-SEP-2000; 2000US-0663561.

PR 20-OCT-2000; 2000US-0693325.

PR 30-NOV-2000; 2000US-0728422.

XX (HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Gao Y, Ma Y,

PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX WPI; 2001-476283/51.

DR P-PSDB; AAM80096.

XX

PT Nucleic acids encoding polypeptides with cytokine-like activities.
PT useful in diagnosis and gene therapy -
XX
XX Claim 1; Page 4968; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78333-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 1164 BP; 273 A; 280 C; 325 G; 286 T; 0 other;

Query Match 27 5%; Score 368.4; DB 22; Length 1164;

Best Local Similarity 84.9%; Pred. NO. 9e-88;

Matches 601; Conservative 0; Mismatches 86; Indels 21; Gaps 16;

Oy 1 GCGGCGCGCCCTGACACAAATGCTCAG-CTTAATGCTCAGCGAGTTCGCTCAACCCAG 59

Db 827 GCGGCGCGCCCTGACACAAATGCTCAGCTTACGCTTCGCGGCGAGTTGCTCCCGCGAG 768

Oy 60 AATGCAATCCT-GCAGAAATACAGGCGCCCTCATTCCCATCCCGCGCGAG-ACACCGGC 117

Db 767 AATGCAATCCTCCCGGGAATACAGGCGCCCTCATTCCCATCCCGCGCGAGTACACAGGC 708

Oy 118 CAGGCGCGCTGCTCCCG-CCACACATTAATTTG-ATGCTCTTACAGAGGAGCTCGGAG 175

Db 707 CAGGCGCGCTGCTCCCGAGACACATTAATTTGATCTTCCCGCGCGAGGAGCTCGGAG 648

Oy 176 CAGAGC---GCTTATACAGGCGACAG-CCGTCTCCGCGA-CCGCGACACAGACAGATGA 230

Db 647 CAGAGCCCGGGGACAGACAGCGCTCAGACCGTCTCTGCGACCGCGACAAACAGATGA 588

Oy 231 TGCGGCGCGCGACGAGCGGCGCGAGCAGC---AACTTCTGAAAAACAGAAAAACA 287

Db 587 CGGAGCAGCCGAGCGATGCGCAAGCCCGACAGACAACTTCTGAAAAACAGAAAAACA 528

Oy 288 GTC-CCAGGCCAAGC-GGCTGCAATGTGTCCAATCCCTTC-CCGTTCCGGGATCCAGA 344

Db 527 GTCCTCAGCCCAAGCGGGCTGCAATGTCTCCAATATCCCTTCAGGGTCCGGGATCCGGA 468

Oy 345 CTTCCGACAAATG-TTGGCGCAATTTGGTAAATATTAGATGTTGAATTTTATG 403

Db 467 CTTCCGACAAATGTTTGGTCAATTTGGTAAATCTTAAATGTTGAATTTTATG 408

Oy 404 AGCGGGGCTGAAAGGATTTGTTGTAACCTTTCGAAAAATAGTCGATGCGACAGCG 463

Db 407 AGCGGGGCTGAAAGGATTTGTTGTAACCTTTCGAAAAATAGTCGATGCGACAGCG 348

Oy 464 CGAGGAGAAA-TTGCACGGTACCGTGG-TAGAGGGCGGCTGAAATGAGTTAA 519

Db 347 CGAGGAGAAA-TTGCACGGTACCGTGGTGAAGGGCGGTAATTTGAGTTAAAT 288

Oy 520 GCGACAGCAGCGTATGATTAATAAAGGCGTGAACCCCTACACCAATGCTGAGAA 579

Db 287 GCGACAGCAGCGTATGATTAATAAAGGCGTGAACCCCTACACCAATGCTGAGAA 228

Oy 580 TTAATCACTTGTGGCGGGCTACAGCCCGGACTTCTATGAGGACAGGCTGTTG 639

Db 227 TTGAATCACTTGTGGCTGAGCTACAGTCCGGAATTTATGACAGGACAGGCTGTTG 168

Oy 640 TGCCAGGCAACAGAGGAGATTTCCATGTACAGTGGGCCCGAGTTCA 687

Db 167 TGCCAGGCAACAGAGGAGATTTCCATGTACAGTGGGCCCGAGTTCA 120

DE Human cDNA sequence SEQ ID NO:10786.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakematsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 8; SEQ ID 10786; 2537bp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification, where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 1513 BP; 354 A; 418 C; 395 G; 346 T; 0 other;
XX
XX Query Match 67.6%; Score 905.2; DB 22; Length 1513;
XX Best Local Similarity 86.7%; Pred. No. 1.4e-230;
XX Matches 1118; Conservative 0; Mismatches 123; Indels 49; Gaps 9;

Qy 176 CAGAGC---GCTTATACAGCGCACAG-CCGCTTCGCGCACCGCCACAGACATGAT 231
Db 443 CAGAGCCCGCGGACAGAGCGCTCAGACCGCTCTGTGCAACCGCCACAGAGCATGAC 502
Qy 232 GCGCCCGCCGACCGGACGCGCCCGCCAGACCAACCTTCTGAAAAACAGAAAACAAGTCC 291
Db 503 GAGGACCGAGATGGCGACGCCCGACACACACCTTCTGAAAAACGAAAAACAACTCT 562
Qy 292 CAGCCCAAGCGGCTGATGTGTCCAAACATCCCTTCGCGTTCGCGGATCCAGACTCCGA 351
Db 563 CAGCCCAAGCGGCTGATGTGTCCAAATATCCCTTCAGGTTCCGCGGATCCGAGACTCAGA 622
Qy 352 CAATGTTGGCCAAATTTGTAATAATATGATGTTGAATTAATTTTATGACGGGGC 411
Db 623 CAATGTTGGTCAATTTGTAATAATCTTAATATTTGAATTAATTTTATGAGGAGAGC 682
Qy 412 TCGAAGGATTTGTTGCTGTAACCTTGGAATAATGTGCGATGTGGGACAGGGCGAGGAG 471
Db 683 TCGAAGGATTTGTTGCTGTAACCTTGGAATAATGTGCGATGTGGGACAGGGCGAGGAG 742
Qy 472 AAATTGACGCGTACCGTGTAGAGGCGCGTAAATCGAGTTAATGACAGACAGC 531
Db 743 AAATTGACGCGCACCGTGTAGAGGCGCGTAAATCGAGTTAATGACAGACAGC 802
Qy 532 GTGATGACTAATAAAAAGCGCGTGAACCTTACACCAATGCTGGAATTAATCACTT 591
Db 803 GTAATGACAAATTAATAAGACCGTCAACCTTATCAATGCTGGAATTAATCACTT 862
Qy 592 GTGGCGCGGCTACAGCCCGGACCTTATGACGAGCAGGCTGTGTGCGAGGCGAAC 651
Db 863 GTGGCGGAGTACAGTCCGGAATTTATGACGAGCAGGCTGTGTGCGAGGCGAAC 922
Qy 652 CAGAGGAGATCTTCATCATGATAGAGGCGCCGATTCATCTGTATATCTTGTGATGCT 711
Db 923 CAGAGGAGATCTTCATCATGATAGAGGCGCCGATTCATCTGTATATCTTGTGATGCA 982
Qy 712 GAGTTTCAATATCCGCGCCGACAGCTGTGACCTGATACGAGGGGCTCAGCTTGAAGC 771
Db 983 GAGTTTCCGATCCAGACGCGCCGCGCTTACCAAGGGGCGCCAGCTTGTGAGAGC 1042
Qy 772 GGTGTGCGACCGTATCAACACCTTCAAGCTGTGGGCGCCGACCCCAATCCGCGC 831
Db 1043 GCGGTGCGACCGTATCAACACCTTCAAGCTGTGGGCGCCGCGCCGACCCCAATCCGCGC 1102
Qy 832 TATGCGAGATAGTATATCAAGAGCAGTGTGCAATTAATGCTACAGGGTGTATC 891
Db 1103 TACGCGGATGTTGTTTACAGAGTGTATTTATG---TGCAGACATTTATGTGTAT 1159
Qy 892 GCTGATACCGCTACGCGCCAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 951
Db 1160 GCTGATACCGCTACGCGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1219
Qy 952 GAGCAGATTTATGCTGCGGACCCCTTACACACACACTTGTCTCCAGCCCGCCAGCTG 1011
Db 1220 GAGCAGATTTATGCTGCGGACCCCTTACACACACACTTGTCTCCAGCCCGCCAGCTG 1279
Qy 1012 GTTGTGCGATGATGCTTTTGGCGCTTGAACCGATGCGCAAGTACAGCATCTGAT 1071
Db 1280 GTTGTGCGATGATGCTTTTGGCGCTTGAACCGATGCGCAAGTACAGCATCTGAT 1339
Qy 1340 GATGTGGGCTGCTTCTTCTTCAATGAGGCTAGTATATACCGAGGGGATACAAACGT 1399
Qy 1132 TTTGCTCATATTAATGATTAACATTAACAAACAAAGCAAAACAAACAAACAA 1191
Db 1400 TTTGCTCATATTAATGATTAACATTAACAAACAAACAAACAAACAAACAA 1430
Qy 1192 AAAAACAACCTTCAATGTGGGAGAGAGAGCTTCCGAGGCGCGAGTGTGCGAC 1251
Db 1431 -----AACCCTCAATGTGGGAGAGAGAGCTTCCGAGGCGCGAGTGTGCGAAT 1482
Qy 1252 ACATGCAATGAGACATCACTTATGCACTC 1281

XX 09-AUG-2001.
 XX 05-FEB-2001; 2001WO-US04098.
 XX 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Drmanac RT, Arundt V, Zhou P, Xu C, Gao Y, Ma Y;
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejrtman T, Goodrich R;
 XX WPI; 2001-476283/51.
 DR P-PSDB; AAM79112.
 XX Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 PS Claim 1; Page 2643-2645; 6221pp; English.
 XX The invention relates to polynucleotides (AAK51456-AAK51435) and the
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 CC
 XX Sequence 1800 BP; 432 A; 552 C; 481 G; 335 T; 0 other;
 SQ
 Query Match 67.7%; Score 906.8; DB 22; Length 1800;
 Best Local Similarity 86.7%; Pred. No. 5.7e-231;
 Matches 1119; Conservative 0; Mismatches 122; Indels 49; Gaps 9;

QY 352 CAAATGTTGGCCCAATTGGTAAATTTAGATGTAAATTTATTTTAAATGACGGGCGC 411
 DB 910 CAAATGTTGGTCAATTTGGTAAATTTAGATGTAAATTTATTTTAAATGACGGGCGC 969
 QY 412 TCGAAGGAGATTGGTTCGTAACTTTGAAATATGTCGGATGCGACAGGCGAGGAG 471
 DB 970 TCAAGGAGATTGGTTCGTAACTTTGAAATATGTCGGATGCGACAGGCGAGGAG 1029
 QY 472 AAATTCACGGTACCGTGTAGAGGCGCGTAAATTCAGGTTAAATTCGACACGACGC 531
 DB 1030 AAATTCACGGGACCGTGTAGAGGCGCGTAAATTCAGGTTAAATTCGACACGACGT 1089
 QY 532 GTATGATTAATAAAGCCCGTGAACCCCTACCAATGCTGGAATTAATTCAGTT 591
 DB 1090 GTATGATTAATAAAGCCCGTGAACCCCTACCAATGCTGGAATTAATTCAGTT 1149
 QY 592 GTGGGCGCGGTCTACAGCCCGGACCTTATGACGGGCGGCTGTGTCGACGGCCAC 651
 DB 1150 GTGGGTCAGTCTACAGTCCGGAATTTATGACGGGCGGCTGTGTCGACGGCCAC 1209
 QY 652 CAGAGGGATCTTCATGATGACAGTGGCCCGAGTTCACTTGTATATACCTTGCAGTCC 711
 DB 1210 CAGAGGGATCTTCATGATGACAGTGGCCCGAGTTCACTTGTATATACCTTGCAGTCC 1269
 QY 712 GCGTTTCATATCCGCGCCGACCTGCTGACAGTGCATACGAGGGGCTCACCTTCAGGC 771
 DB 1270 GCGTTTCATATCCGCGCCGACCTGCTGACAGTGCATACGAGGGGCTCACCTTCAGGC 1329
 QY 772 CGGTGTGACCCGTTTCAACACCTTGAAGCTGCGGCGCCCGACCCCAATCGCGGCG 831
 DB 1330 CGGTGTGACCCGTTTCAACACCTTGAAGCTGCGGCGCCCGACCCCAATCGCGGCG 1389
 QY 832 TATGCGGAGTAGTATCAAGAGCCAGTATGCAATTAATTTCTACAGGGTGTAC 891
 DB 1390 TATGCGGAGTAGTATCAAGAGCCAGTATGCAATTAATTTCTACAGGGTGTAC 1446
 QY 892 GCTGATACCGCTACGCGCCAGCCACCTGTCAGTGTCTGCTACAGTACAGTTAC 951
 DB 1447 GCTGATACCGCTACGCGCCAGCCACCTGTCAGTGTCTGCTACAGTACAGTTAC 1506
 QY 952 GGAGGATTTATGTCGCGACCCCTACACACACATTTCTCCAGGCCACCTACAGGC 1011
 DB 1507 GGAGGATTTATGTCGCGACCCCTACACACACATTTCTCCAGGCCACCTACAGGC 1566
 QY 1012 GTTGTCATGATGATGTTTGGCGCCCTTGAACGATGCAAGACTAGAGCCATGTCAT 1071
 DB 1567 GTTGTCATGATGATGTTTGGCGCCCTTGAACGATGCAAGACTAGAGCCATGTCAT 1626
 QY 1072 GATGTGGGTCGCTCTTCTTCTTCAATGACGCTATATATACCAAGGGGATACACCGT 1131
 DB 1627 GATGTGGGTCGCTCTTCTTCTTCAATGACGCTATATATACCAAGGGGATACACCGT 1686
 QY 1132 TTTGCTCCATTTAAATGATTAACATTTAAACAAACCAACCAAAAAACAAAAACA 1191
 DB 1687 TTTGCTCCATTTAAATGATTAACATTTAAACCAACCAAAAAACCAAAAAACA 1217
 QY 1192 AAAAAACCACTTCCATGTGGGAGAGAGAAAGTTTCCGAGGCCGAGTGTGGCAG 1251
 DB 1718 -----AACCTTCCAAATGTGGGAGAGAAAGCTTTCCGAGGCCGAGTGTGGCAGT 1769
 QY 1252 ACATGACGTAGGACATCATTTAGCAATC 1281
 DB 1770 ACATGACGTAGGACATCATTTAGCAATC 1799

RESULT 3
 AAH13824
 ID AAH13824 standard; cDNA, 1513 BP.
 XX AAH13824;
 AC
 XX
 XX 26-JUN-2001 (first entry)
 DT
 XX

PA (SCIO-) SCIOS INC.
 XX Stanion LM, White RT;
 XX WPI: 2002-010779/01.
 DR P-PSDB; AAU070146.
 XX
 PT Novel secreted factor polypeptide useful for treating cardiac diseases
 PT such as arteriosclerosis, myocardial infarction, inflammatory diseases
 PT such as asthma, stroke, and rheumatoid arthritis and renal diseases -
 XX
 PS Claim 1; Fig 1; 189pp; English.
 XX
 CC The invention relates to rat secreted factor polypeptides and the
 CC polynucleotides encoding them. The sequences are useful for treating
 CC cardiac, renal or inflammatory diseases. These include cardiac diseases
 CC such as congestive heart failure, myocarditis, dilated congestive
 CC cardiomyopathy, angina pectoris, myocardial infarction, cardiac
 CC arrhythmia, pulmonary hypertension, arteriosclerosis, atherosclerosis and
 CC cardiac tumours, renal diseases such as glomerulonephritis, nephrotic
 CC syndrome, renal infarction, hereditary nephritis, polycystic kidney
 CC disease, chronic renal failure, renal vein thrombosis and medullary
 CC sponge kidney and inflammatory diseases such as asthma, rheumatoid
 CC arthritis, osteoarthritis, stroke, psoriasis, resenosis, graft versus
 CC host reaction, Crohn's disease, ulcerative colitis and Alzheimer's
 CC disease. Sequences AAS94693-AAS94745 represent cDNA clones, which encode
 CC the secreted factor polypeptides of the invention, and oligonucleotide
 CC probes and PCR primers.
 CC
 XX
 SQ Sequence 1340 BP; 366 A; 394 C; 317 G; 263 T; 0 other;
 Query Match 100.0%; Score 1340; DB 24; Length 1340;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 541 AATAAAGGCGGTGAACCCCTACACCAATGCTGGAATTAATTAATCCAGTTGTGGCGG 600
 Qy GTCTACAGCCCGGACTTCTATGACAGGACGGGTGTTGTGTCAGGCCAACAGAGGGA 660
 Db 601 GTCTACAGCCCGGACTTCTATGACAGGACGGGTGTTGTGTCAGGCCAACAGAGGGA 660
 Qy 661 TCTTCATGTAAGTGGGCCCGGACTTCACTGTATATACCTTGTGCAATGCTGGCTTTCA 720
 Db 661 TCTTCATGTAAGTGGGCCCGGACTTCACTGTATATACCTTGTGCAATGCTGGCTTTCA 720
 Qy 721 TATCCGGCGGCGACTGCTGACGCTGCAATACGAGGGGCTCACTTTCAGGCGGTGTGC 780
 Db 721 TATCCGGCGGCGACTGCTGACGCTGCAATACGAGGGGCTCACTTTCAGGCGGTGTGC 780
 Qy 781 ACCGTGTACACACCTTACAGAGTGGCGCGGCCACCCCAATCCGCGCTATGCGGA 840
 Db 781 ACCGTGTACACACCTTACAGAGTGGCGCGGCCACCCCAATCCGCGCTATGCGGA 840
 Qy 841 GTAGTGTATCAAGAGCCAGTATGCAATTAATTTGCTACAGGGGTGTTACGTCATAC 900
 Db 841 GTAGTGTATCAAGAGCCAGTATGCAATTAATTTGCTACAGGGGTGTTACGTCATAC 900
 Qy 901 CGTACGCGCCAGCCCAACCCCTGCGCACTGCTGCTACAGTGAAGTTAGGACGAGTT 960
 Db 901 CGTACGCGCCAGCCCAACCCCTGCGCACTGCTGCTACAGTGAAGTTAGGACGAGTT 960
 Qy 961 TATGCTGCGACCCCTTACACCACTTGTCTCAGGCCCCACCTACGCGGTGTGCC 1020
 Db 961 TATGCTGCGACCCCTTACACCACTTGTCTCAGGCCCCACCTACGCGGTGTGCC 1020
 Qy 1021 ATGAATGCTTTTGGCGCCCTTACCGATGCAAGACTAGAGCCATGCTGATGATGGGT 1080
 Db 1021 ATGAATGCTTTTGGCGCCCTTACCGATGCAAGACTAGAGCCATGCTGATGATGGGT 1080
 Qy 1081 CTGCTTTCTTTCTTATGACAGCTAGTATATACCAAGGGGATCAACGCTTTGCTCA 1140
 Db 1081 CTGCTTTCTTTCTTATGACAGCTAGTATATACCAAGGGGATCAACGCTTTGCTCA 1140
 Qy 1141 TATTAAATGATTAACCACTTAAACAAACAAACAAACAAACAAACAAACAAACAA 1200
 Db 1141 TATTAAATGATTAACCACTTAAACAAACAAACAAACAAACAAACAAACAAACAA 1200
 Qy 1201 ACCTTCAATGTGGGAGAGAGAAAGCTTTCCGAGGCCGAGTGTGCGACATGCACT 1260
 Db 1201 ACCTTCAATGTGGGAGAGAGAAAGCTTTCCGAGGCCGAGTGTGCGACATGCACT 1260
 Qy 1261 AGGACATCACTTTAGCACTCAAGAAACAAACGAAAAAATTAAGC 1320
 Db 1261 AGGACATCACTTTAGCACTCAAGAAACAAACGAAAAAATTAAGC 1320
 Qy 1321 GGCCGAGGGGTTGCTAGA 1340
 Db 1321 GGCCGAGGGGTTGCTAGA 1340

RESULT 2
 ID AAK52245 standard; cDNA; 1800 BP.
 XX AAK52245;
 AC 06-NOV-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 790.
 XX
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukemia;
 KW nervous system disorder; arthritis; inflammation; ss.
 OS Homo sapiens.
 XX
 PN W0200157190-A2.

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 10:24:02 ; Search time 314.202 Seconds
(without alignments)
9604.255 Million cell updates/sec

Title: US-09-809-545A-1

Perfect score: 1340
Sequence: 1 gcggcgcccccgcacacacat.....ggccgaagggttcgctaga 1340

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: N_GenSeq_101002.*
2: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1980.DAT.*
3: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1981.DAT.*
4: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1982.DAT.*
5: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1983.DAT.*
6: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1984.DAT.*
7: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1985.DAT.*
8: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1986.DAT.*
9: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1987.DAT.*
10: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1988.DAT.*
11: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1989.DAT.*
12: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1990.DAT.*
13: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1991.DAT.*
14: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1992.DAT.*
15: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1993.DAT.*
16: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1994.DAT.*
17: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1995.DAT.*
18: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1996.DAT.*
19: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1997.DAT.*
20: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1998.DAT.*
21: /SID52/gcgdata/geneeq/geneeqn-emb1/NA1999.DAT.*
22: /SID52/gcgdata/geneeq/geneeqn-emb1/NA2000.DAT.*
23: /SID52/gcgdata/geneeq/geneeqn-emb1/NA2001A.DAT.*
24: /SID52/gcgdata/geneeq/geneeqn-emb1/NA2001B.DAT.*
25: /SID52/gcgdata/geneeq/geneeqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1340	100.0	1340	24	AA594693
2	906.8	67.7	1800	22	AAK52245
3	905.2	67.6	1513	22	AAH13824
4	857.2	64.0	2372	21	AAH07075
5	385.2	28.7	1295	22	AA531173
6	368.4	27.5	1164	22	AAK53129
7	338.8	25.3	607	22	AAH09205
8	334	24.9	893	22	AAH03678
9	319	23.8	481	23	AA593633

C	10	290	21.6	539	24	ABK34530	Human CDNA for nov
C	11	286.6	21.4	939	22	AA161034	Human polynucleoti
C	12	286.6	21.4	939	22	AA161035	Human polynucleoti
C	13	286.6	21.4	1011	22	AA159248	Human polynucleoti
C	14	275.6	20.6	1506	21	AA246827	Human RNA binding
C	15	251.8	18.8	327	22	AAK54001	Murine transcripti
C	16	229	17.1	406	23	AA593532	DNA encoding novel
C	17	170.6	12.7	3189	24	ABK34602	Human CDNA for nov
C	18	154	11.5	395	22	ABH11642	Human nervous syst
C	19	146	10.9	903	23	ABH13665	Drosophila melanog
C	20	145.4	10.6	394	20	AA566245	DNA encoding novel
C	21	141.8	10.6	394	20	AAV87007	EST clone AY147.
C	22	134.4	10.0	918	22	AA59249	Human polynucleoti
C	23	123.8	9.2	473	22	AAK07716	Human brain expres
C	24	123.8	9.2	473	22	AAK33549	Human bone marrow
C	25	123.8	9.2	473	22	AA139304	Probe #17990 used t
C	26	123.8	9.2	473	24	AB508416	Human genome-deriv
C	27	109.6	8.2	125	22	AAK20380	Human brain expres
C	28	109.6	8.2	125	22	AAK46486	Human bone marrow
C	29	109.6	8.2	125	22	AA152354	Probe #21040 used
C	30	109.6	8.2	125	24	AB520830	Human genome-deriv
C	31	107	8.0	2118	23	AA593634	DNA encoding novel
C	32	93	6.9	1252	23	AA590510	DNA encoding novel
C	33	88	6.6	224	22	ABH70887	Human foetal liver
C	34	88	6.6	224	22	ABH37333	Probe #15799 for g
C	35	88	6.6	224	22	AAK19155	Human brain expres
C	36	88	6.6	224	22	AAK45114	Human bone marrow
C	37	88	6.6	224	22	AA125142	Probe #15075 for g
C	38	88	6.6	224	22	AA151072	Probe #19758 used
C	39	88	6.6	224	24	ABH519164	Human genome-deriv
C	40	86.4	6.4	454	22	ABH58290	Human foetal liver
C	41	86.4	6.4	454	22	ABH27427	Probe #5893 for ge
C	42	86.4	6.4	454	22	AAK06386	Human brain expres
C	43	86.4	6.4	454	22	AAK32056	Human bone marrow
C	44	86.4	6.4	454	22	AA115924	Probe #5857 for ge
C	45	86.4	6.4	454	22	AA137912	Probe #6598 used t

ALIGNMENTS

RESULT 1	AA594693	standard; CDNA; 1340 BP.
XX	AA594693:	
AC	12-MAR-2002.	(first entry)
XX		
DT		
XX		
DE		Rat secreted factor DNA clone P0184_D11 #1.
XX		
KW		Rat; secreted factor polypeptide; cardiac disease; renal disease; kidney;
KW		inflammatory disease; congestive heart failure; myocarditis; asthma; ss;
KW		dilated congestive cardiomyopathy; angina pectoris; cardiac arrhythmia;
KW		myocardial infarction; pulmonary hypertension; arteriosclerosis; stroke;
KW		atherosclerosis; cardiac tumour; glomerulonephritis; nephrotic syndrome;
KW		renal infarction; hereditary nephritis; polycystic kidney disease;
KW		chronic renal failure; renal vein thrombosis; medullary sponge kidney;
KW		rheumatoid arthritis; osteoarthritis; psoriasis; restenosis; PCR primer;
KW		graft versus host reaction; Crohn's disease; ulcerative colitis; probe;
XX		Alzheimer's disease; gene therapy.
XX		
OS		Rattus norvegicus.
XX		
PN		WO200174901-A2.
XX		
PD		11-OCT-2001.
XX		
XX		
PF		23-MAR-2001; 2001WO-US09555.
XX		
PR		31-MAR-2000; 2000US-193548P.
XX		
PR		14-MAR-2001; 2001US-0809545.
XX		

REFERENCE
AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 115860)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alabrooks, S.L., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T.,
Barbaria, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlssoon, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulsegod, H.,
Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenwo, S., Ogih, M., Okwuonu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojokoban, I., Rolfe, M., Ruiz, S., Savary, G.,
Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,
Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Uman, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Woodson, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 115860)
Worley, K.C.
Direct Submission
Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 115860)
Worley, K.C.
Direct Submission
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 18, 2002 this sequence version replaced gi:20514522.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GXLC
Center clone name: CH230-34C2
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 99693 bases at least Q40
Consensus quality: 101936 bases at least Q30
Consensus quality: 102743 bases at least Q20

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNALREFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 1347: contig of 1347 bp in length
* 1348 1447: gap of unknown length
* 1448 2834: contig of 1387 bp in length
* 2835 4253: contig of 1319 bp in length
* 4254 4353: gap of unknown length
* 4354 5477: contig of 1124 bp in length
* 5478 7484: contig of 1907 bp in length
* 7485 7584: gap of unknown length
* 7585 9406: contig of 1822 bp in length
* 9407 9507: gap of unknown length
* 9508 11712: contig of 2206 bp in length
* 11713 13462: contig of 1650 bp in length
* 13463 13562: gap of unknown length
* 13563 17222: contig of 3866 bp in length
* 17223 20617: gap of unknown length
* 20618 24825: contig of 4108 bp in length
* 24826 28911: contig of 3886 bp in length
* 28912 33686: contig of 4775 bp in length
* 33687 40217: contig of 6431 bp in length
* 40218 47912: contig of 7595 bp in length
* 47913 48012: gap of unknown length
* 48013 56482: contig of 8470 bp in length
* 56483 56582: gap of unknown length
* 56583 66204: contig of 9622 bp in length
* 66205 74546: contig of 8242 bp in length
* 74547 85222: contig of 11876 bp in length
* 85223 86623: gap of unknown length
* 86624 100049: contig of 13427 bp in length
* 100050 100149: gap of unknown length
* 100150 115860: contig of 15711 bp in length.
FEATURES
Location/Qualifiers
1..115860
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-34C2"
BASE COUNT 33039 a 22417 c 22786 g 30574 t 7044 others
ORIGIN
Query Match 21.3%; Score 285.4; DB 2; Length 115860;
Best Local Similarity 98.0%; Pred. No. 5.8e-66;
Matches 289; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 1023 GAATGCTTTTGGCCCTGACCGATGCCAGCTAGAGCCATCTGATGATGGGTCT 1082
Db 112555 GAATGCTTTTGGCCCTGACCGATGCCAGCTAGAGCCATCTGATGATGGGTCT 112496
Qy 1083 CGTTCTTTTCTTCATTGTCAGGCTAGTATATACCAAGGGGGATACAAACCGTTTGGTCCATA 1142
Db 112495 CGTTCTTTTCTTCATTGTCAGGCTAGTATATACCGAGGGGGATACAAACCGTTTGGTCCATA 112436
Qy 1143 TTAATGATATAAACCATTAACAAACCAAGCAAAAAACAAACCAAAACCAAC 1202

Query Match		40.4%; Score 541.2; DB 9; Length 1623;	
Best Local Similarity		89.0%; Pred. No. 6.7e-136;	
Matches 664; Conservative		0; Mismatches 73; Indels 9; Gaps 7;	
Qy	1	GGGGCGGCGCCCTGACAAATGGCTCAG-CTTATGCCTCAGCGAGTTCGCTCCACCCGAG	59
Db	878	GGCGTGCCCTTGACAAATGGCTCAGCTTACGCTTCGGCCAGTTTGTCTCCCGCGAG	937
Qy	60	AATGGCATCT-GCAGAATACAGCGGCCCTCATCCCCATCCCGGCCACAGAG-ACACCGGC	117
Db	938	AACGGTATCCCGCGGAATACAGCGCCCTCATCCCAACCCGCGCAGATACACAGCG	997
Qy	118	CAGCCCATGTCCCG-CCACACATTAACCTTG-ATCCTCTACACAGACGCACTCCGAG	175
Db	998	CAGACACCGTTCGCGAGCAGACATTAACCTGTACCTCCCGCCACAGCACTCCGAG	1057
Qy	176	CAGAGC---GCTTATACAGGCGACAG-CGCTCTCCGCGCACCGCCACACAGATGAT	231
Db	1058	CAGAGCCGCGGACACAGAGCGCTCAGACCGTCTCTGGCACCGCCACACAGAGTAC	1117
Qy	232	GCGGCGCCGACGAGCGGCCAGACCACTTCTGAAAAACACAGAAAAACAAGTCC	291
Db	1118	GCAGCACCGAGTGGCGACGCCCCAGACAACTTCTGAAAAACACGGAACAACTCT	1177
Qy	292	CAGCCCAAGCGGCTGCATGTCTCAACATCCCTTCGGTTCGGGATCCAGACTCCGA	351
Db	1178	CAGCCCAAGCGGCTGCATGTCTCANAATCCCTTCAGGTTCCGGATCCGACCTCAGA	1237
Qy	352	CAAAATGTTGGCCAAATTTGGTAAATATTAGATGTGAAATATTATTAATGAGCGGGC	411
Db	1238	CAAAATGTTTGGTCAATTTGGTAAATCTTAGATGTGAAATATTATTAATGAGCGAGC	1297
Qy	412	TCGAAGGATTTGGTTTGGTAACTTTTGGAAATAGTCGGATCGGACAGCGCGAGGAG	471
Db	1298	TCAAAGGATTTGGTTTGGTAACTTTTGGAAATAGTCGGATCGGACAGCGCGAGGAG	1357
Qy	472	AAATTTGACCGTACCGTGTAGAGGCGGTAAATTCAGGTTAATAATGCGACAGCAGC	531
Db	1358	AAATTTACGCGCACCGTGTAGAGGCGGTAAATTCAGGTTAATAATGCGACAGCAGT	1417
Qy	532	GTGATGACTAATAAAGCGGTGAACCCCTACACCAATGGCTGGAATTTAAATCCAGTT	591
Db	1418	GTAATGACAAATAAAGACCGTCAACCCCTATACAAATGGCTGGAATTTGAATCCAGTT	1477
Qy	592	GTGGGCGGCTCTACAGCCCGGACTTCTATGACGACGCGTCTGTGTGCGAGCGCAAC	651
Db	1478	GTGGGTGCAGTCTACAGTCCCGAATCTATGACGACGCGTCTGTGTGCGAGCGCAAC	1537
Qy	652	CAGGAGGATCTTCATGTACAGTGCGCCCGAGTTCATTTGTATATACCTTCTGCAATGCC	711
Db	1538	CAGGAGGATCTTCATGTACAGTGCGCCCGAGTTCATTTGTATATACCTTCTGCAATGCCA	1597
Qy	712	GGCTTCCATATCCGCGCGCACTGC	737
Db	1598	GGCTTCCGATATCAGCAGCCACCGC	1623
RESULT 13			
AF387322			
LOCUS			
DEFINITION			
AF387322			
ACCESSION			
AF387322.1			
VERSION			
GI:14495355			
KEYWORDS			
SOURCE			
ORGANISM			
Mus musculus.			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
1 (bases 1 to 1134)			
Lieberman,A.P., Friedlich,D.L., Harmison,G., Howell,B.W.,			
Jordan,C.L., Breedlove,S.M. and Fischbeck,K.H.			
Androgens regulate the mammalian homologues of invertebrate sex			
determination genes tra-2 and fox-1			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
1..1134			
/organism="Mus musculus"			
/db_xref="taxon:10090"			
1..1134			
/note="similar to Caenorhabditis elegans fox-1; contains			
RSM domain"			
/codon_start=1			
/product="putative RNA-binding protein ftx"			
/protein_id="AAK64287.1"			
/db_xref="GI:14495356"			
/translation="MEKKMTVGNOEPTTTPDMVOPFTTIPPPPPQNGIPTBYGV			
PHTQYAGOTSEHNLTLYGSTQPHGEQSSNPSNGSLTQEGGAQDQGOQSOTSS			
ENSEKSTPKRLHVSNIPIFRPDPLRMFGQFGKILDVEIIFNERSKGFGVTFPEN			
SADARAREKLHGTVEGRKIEVNATARVMTKMTVPYANGKLSVVGAVYGPFL			
YAASSFOADVSLGNEAAVPLSGRGINTYIPLIIPGFPYTAATAAFAFRGHLRGRG			
RTVYAVRAVPPTAIPAYGVYODGYADLYGGAAYRYAQAPATATAATAAAAAA			
AYSDGGRVYTADPVHALAPAAASYGVGAVASLYRGYSRPAPY"			
BASE COUNT			
ORIGIN			
301 a 312 c 293 g 228 t			
Query Match			
Best Local Similarity			
Matches 587; Conservative			
0; Mismatches 268; Indels 51; Gaps 8;			
Qy	160	ACAGACGCACTCGGAGCAGAGCGCTTATACAGCGCACAGCGCTCTCCGGCACCCGCCACA	219
Db	193	ACGCGACCTCATGGAGAACAGAGTAGCAATTCACCCAGCAACCAAGATGGATCTCTCAGC	252
Qy	220	CAGACAGATGATGCCCGCCGACCGACCGG---CCAGCCCCGACACACACCTTCTGAAAAAC	276
Db	253	CAGACAGAGGTGGAGCAGACAGACAGCAGCGGACCAACAGTTCACAGACACAAAGTAGTGAATAAT	312
Qy	277	ACAGAAAAACAAGTCCAGCCCCAAGCGGTGCATGTGTCCAAATCCCCCTTCGGGTTCGGG	336
Db	313	TCAGAGAGTAAATCTACGCCCAAGCGACTACATGTCTCTAATATTTCCCTTCGGCTTCGGA	372
Qy	337	GATCCAGACCTCCGACAAATGTTTGGCCAAATTTGGTAAATAATTAGATGTTGAAATTAAT	396
Db	373	GACCTGACCTCCGCGCAGATGTTTGGGCAGTTTGGCAAAATCCTAGATGTGGAATAATC	432
Qy	397	TTTAATGAGCGGGCTCGAAGGGATTTGGTTTTCGTAACCTTTTCAAAATAGTCCGATGCG	456
Db	433	TTTAATGAGCGGGCTCCAGGGATTCGGGTTCGTAACTTTTCGAGATAGTGTCTGATGCA	492
Qy	457	GACAGGCGAGGGAGAAATTCACGGTACCGTGTAGAGGCCCGTAAATCGAGGTTAAT	516
Db	493	GACAGGCGAGGGAGAAATTCACGGCACCGCTGTAGAGGCCCGTAAATCGAGGTTAAT	552
Qy	517	AATGCGACAGCAGCGTGATGACTAATAAAGCGCGTGAACCCCTACACCAATGGCTGG	576
Db	553	AATGCAACAGCAGCGGTCTATGCAACAGAGATGGTCAGCCCATATGCAAAATGGCTGG	612
Qy	577	AAATTAATCAGTGTGGGCGCGGTCTACAGCCCCGACTTCTATGAGGCGACCGTGTCT-	635
Db	613	AAGTTAAGCCCAAGTAGTTGGAGCTGTGTATGGTCTCAGTTATATGAGCATCAGCTTT	672
Qy	636	-----GTTGTGCCAGGCCAACACAGGAGGAGATCTTCCATGTACAGTGCGCCAGTTCA	687
Db	673	CAAGCTGATGTGCTCCCTAGGCAATGAGCGGGCTGTGCCCTTGTTCAGGAAGAGGAGCATC	732
Qy	688	CTTGATATACCTCTGCA---ATGCGCTGGCTTTCCATATC---CGGCCGCACTGCTGCA	741
Db	733	AACACTTACATCCCTCTAATCATCTTCTGGCTTCCCTTACCAACTTCAGCAGCAGCAGGCA	792

```
/product="hexaribonucleotide binding protein 1 isoform
alpha"
/protein_id="AAL83405.1"
/db_xref="GI:19032366"
/translation="MLASQGVLLHPYGVPMIVPAAPYLPGLIQNQEAAPDPTMAQP
YASQAPAPONGIPAEYTPAPHPHAPPEYTGQTPPEHTLNLXPAPQTHSEQSPADTSA
QTVSGTATOTDDAAPDGPOTOPSTENKSPKRLHVSNTIPFRDPLDRQMFQOF
KILDVIELFNERGSKRGFVTEENSADADAREKLGHTVVEGRKLEVNATARVMTN
KKTVPYTNKLNPPVGVAVYSPFAGTAVLLCOANQEGSSMYSAPESSIVVTSAMPGF
PYPAATAAARAGAHURGRRTVINFRAPAAPPPPIPAYGGVYQBPVYGNKLLOGGY
AYRYAQPTPATAYSDSYGRVYAADPVHAPAPTYGVGAMASIIYRGGYRNFAPY
"
BASE COUNT      380 a   448 c   374 g   345 t
ORIGIN
Query Match      59.6%; Score 798.4; DB 9; Length 1547;
Best Local Similarity 89.9%; Pred. No. 7e-206;
Matches 935; Conservative 0; Mismatches 96; Indels 9; Gaps 7;
Qy 1 GCGGCGCCCTGCACAAATGCTCAG-CTTATGCTCAGCGCAGTTCGCTCCACCCCGAG 59
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 356 GCGGCTGCGCCCTGCACAAATGCTCAGCTTACGCTTCGGCCCGAGTTGCTCCCGCGAG 415
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 60 AATGCAATCTT-GCAGAATACAGCGCCCTCATCCCATCCCGCGCCAGAG-ACACCGGC 117
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 416 AACGGTATCCCGCGGATACAGCGCCCTCATCCCATCCCGCGCCAGAGTACACAGGC 475
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 118 CAGCCCATGTCCTCGG-CGACATTAACCTTG-ATCCTCTTACACAGCGCAGCTCCGGAG 175
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 476 CAGACCGGGTTCGCGAGCACACATTAACCTGTACCTCCCGCGCCAGCGCAGCTCCCGAG 535
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 176 CAGAGC---GCTTATACAGCGCACAG-CGGTCTCCGCGCAGCGCCACACAGACGATGAT 231
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 536 CAGAGCGCGCGGACAGAGCGCTCAGACCGTCTCTGGCAGCGCCACACAGACGATGAC 595
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 232 GCGGCGCCCGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 471
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 596 GCAGCAGCGCGGAGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 655
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 292 CAGCCCAAGCGGTGATGTGTCACATCCCTCCGCTCCGCGGATCCAGACCTCCGA 351
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 656 CAGCCCAAGCGGTGATGTGTCACATCCCTCCGCTCCGCGGATCCAGACCTCCGA 715
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 352 CAAATGTTGGCAATTTGCTAAATATTAGATGTTGAAATATTATTTTAAATGAGCGGGC 411
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 716 CAAATGTTGGCAATTTGCTAAATATTAGATGTTGAAATATTATTTTAAATGAGCGGGC 775
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 412 TCAGAGGGATTTGGTTTCGTAATTTTCGAAATATGTCGGATCGGACAGGCGGAGGAG 471
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 776 TCAGAGGGATTTGGTTTCGTAATTTTCGAAATATGTCGGATCGGACAGGCGGAGGAG 835
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 472 AAATTGCAAGTACCGTGTAGAGCGCGCTTAAATGAGGTTAATTAATCGACAGCAGC 531
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 836 AAATTGCAAGTACCGTGTAGAGCGCGCTTAAATGAGGTTAATTAATCGACAGCAGC 895
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 532 GTGATGACTAATAAAGGCGGTGAACCCCTACACCAATGCGTGGAAATTAATTCAGTT 591
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 896 GTAATGCAATTAAGAAGACCGTCAACCTTATACAAATGCGTGGAAATTAATTCAGTT 955
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 592 GTGGGCGCGTCTACAGCCCGGCTTCTATGCGAGGACGCGTGTGTTGCGGCGGCAAC 651
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 956 GTGGGTCAGTCTACAGTCCGAAATCTATGCGAGGACGCGTCTGTTGTCGCGGCAAC 1015
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 652 CAGGAGGATCTTCATGATACAGTGCGGCCCGGCTTCTGATATATCTTCTGCAATGCCT 711
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1016 CAGGAGGATCTTCATGATACAGTGCGGCCCGGCTTCTGATATATCTTCTGCAATGCCA 1075
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 712 GGCTTTTCATATCGGCGCGGCTGCTGCGAGCTGCATACCGAGGGGCTCACTTCGAGGC 771
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1076 GGCTTCCGATCCAGAGAGCAGCCGCGCGCGCTTACCGAGGGGCGACCTTCGAGGC 1135
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 772 CGTGGTCCGACCGGTGTAACACCTTCAGAGCTCGGCGCGCCACCCCAATCCCGGCC 831
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 1136 CCGGTGCGCACCGTGTACAAACACCTTCAGGCGCGCGCGCGCGCGCGCGCGCGCC 1195
Qy 832 TATGCGGAGTAGTGTATCAAGAGCGCAGTGTATGCAATAAATTTCTACAGGGTGTAC 891
Db 1196 TACGCGGAGTAGTGTATCAAGAGCGCTGTGTATGCAATAAATTTCTGCGAGGTTAC 1255
Qy 892 GCTGATACCGTACGCCAGCCACCCCTGCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCT 951
Db 1256 GCTGATACCGTACGCCAGCGCTACCCCTGCGCAGCTGCTGCTGCTGCTGCTGCTGCT 1315
Qy 952 GGACGAGTTTATGCTGCGCAGCCCTTACACACACACTTGTCTCCAGCCCCCCTACGCG 1011
Db 1316 GGACGAGTTTATGCTGCGCAGCCCTTACACACACACTTGTCTCCAGCCCCCCTACGCG 1375
Qy 1012 GTTGGTGCCATGAATGCTTT 1031
Db 1376 GTTGGTGCCATGGCTAGTAT 1395

RESULT 12
HSM803066      1623 bp      mRNA      linear      PRI 20-MAR-2002
LOCUS          Homo sapiens mRNA; cDNA DKF2p547L059 (from clone DKF2p547L059).
DEFINITION     AL713700
ACCESSION      AL713700.1 GI:19584415
VERSION        human.
KEYWORDS       SOURCE
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 1623)
AUTHORS        Bloembergen, H., Boecher, M., Brandt, P., Mewes, H.W., Weil, B. and
                Wiemann, S.
TITLE           Direct Submission
JOURNAL         Submitted (12-MAR-2002) MIPS, Am Klopferspitze 18a, D-82152
                Martinsried, GERMANY
COMMENT         Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
                Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
                sequenced by GBF (National Research Centre for Biotechnology Ltd.,
                Braunschweig/Germany) within the cDNA sequencing consortium of the
                German Genome Project.
                This clone (DKF2p547L059) is available at the RZPD in Berlin.
                Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
                Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
                information about the clone and the sequencing project is available
                at http://mips.gsf.de/proj/cDNA/.
                Location/Qualifiers
                1..1623
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /map="16p13.3"
                /clone="DKF2p547L059"
                /tissue_type="brain"
                /clone_lib="547 (synonym: hfbri). Vector pSport1; host
                DH10B; sites NotI + SalI"
                /dev_stage="fetal"
                1..1623
                /gene="DKF2p547L059"
                836..>1623
                /gene="DKF2p547L059"
                /note="FLJ10165 fis"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAD28499.1"
                /db_xref="GI:19584416"
                /translation="MNCERQLRGNQEAAPDPTMAQFYASQAPQNGIPAEYTA
                PHPAPEYTGQTPPEHTLNLXPAPQTHSEQSPADTSAQTVSGTATOTDDAAPDGP
                QTVSGTATOTDDAAPDGPOTOPSTENKSPKRLHVSNTIPFRDPLDRQMFQOF
                KILDVIELFNERGSKRGFVTEENSADADAREKLGHTVVEGRKLEVNATARVMTN
                KKTVPYTNKLNPPVGVAVYSPFAGTAVLLCOANQEGSSMYSAPESSIVVTSAMPGF
                PYPAATAAARAGAHURGRRTVINFRAPAAPPPPIPAYGGVYQBPVYGNKLLOGGY
                AYRYAQPTPATAYSDSYGRVYAADPVHAPAPTYGVGAMASIIYRGGYRNFAPY
                "
                BASE COUNT      356 a   523 c   487 g   257 t
                ORIGIN
```


REFERENCE	1 (bases 1 to 2372)	Score	857.2	DB 6	Length	2372
AUTHORS	Fulst,S.M. and Shibata,H.					
TITLE	Nucleic acids encoding ataxin-2 binding proteins					
JOURNAL	Patent: US 6194171-A 1 27-FEB-2001;					
FEATURES	Location/Qualifiers					
source	1..2372					
BASE COUNT	548 a 726 c 646 g					
ORIGIN	1..2372					
Query Match	64.0%					
Best Local Similarity	82.7%					
Matches 1141	Conservative 0					
Mismatches 138	Indels 101					
Gaps 10						
Qy	1	CGCGCGCGCCCTGACACAATGGCTCAG-CTTATGCCTCAGCGCAGTTCGCTCCACCCGAG	59			
Db	1029	GCGGTGCGCCCTGACACAATGGCTCAGCTTACGCTTCGGCCAGTTTGTCTCCCGCGAG	1088			
Qy	60	AATGCATCCT-GCAGAAATACACGCGCCCTCATCCCATCCCGGCCACAG-ACACCGGC	117			
Db	1089	AACGTATCCCGCGGAATACACGCGCCCTCATCCCAACCCCGCCGACAGTACACAGGC	1148			
Qy	118	CAGCCACTGTCCCG-CCACACATTAACCTG-ATCCTCTACACAGACGCACTCGGAG	175			
Db	1149	CAGACCACGGTTCGCGAGCACATTAACCTGTACCTCCCGCCACAGCACTCCGAG	1208			
Qy	176	CAGAGC---GCTTATACAAGCGCACAG-CCTGCTCCGSCACCGCACACACAGATGAT	231			
Db	1209	CAGAGCCCGGGACACGAGCGCTCAGACCCTCTGCGCACCGGCACACAGAGATGAC	1268			
Qy	232	CCGCGCCGACCGACGCGCCAGCCCGACACAACTTCTGAAAAACACAGAAAACAAGTCC	291			
Db	1269	GCAGACCGACGGATGGCAGCCCGACACAACTTCTGAAAACCGAAAAACAAGTCT	1328			
Qy	292	CAGCCACGCGCTGCATGTGTCAACATCCCTTCGGTTTCGGGATCCAGACCTCCGA	351			
Db	1329	CAGCCACGCGCTGCATGTGTCCAAATCCCTTCAGTTTCGGGATCCGACCTCAGA	1388			
Qy	352	CAAAATGTTTGGCCAAATTTGGTAAAAATATTAGATGTCGAAATATTATTTATGAGCGGGC	411			
Db	1389	CAAAATGTTTGGTCAATTTGGTAAATCTTAGATGTTGAAATATTATTTATGAGCGAGC	1448			
Qy	412	TCGAGGGATTTGGTTTCGTAACCTTTCGAAATATAGTCGGATCGGACAGGGGAGGGAG	471			
Db	1449	TCGAGGGATTTGGTTTCGTAACCTTTCGAAATATAGTCCGATCGGACAGGGGAGGGAG	1508			
Qy	472	AAATTGACGGTACCGTGGTAGAGGCGCGTAAATTCAGGTTTAAATGCGACAGCACGC	531			
Db	1509	AAATTACCGGACCGTGGTAGAGGCGCGTAAATTCAGGTTTAAATGCGACAGCACGC	1568			
Qy	532	GTGATGACTAATAAAGGCGGTGAACCCCTACACCAATGGCTGGAAATTAATTCAGTT	591			
Db	1569	GTAATGACAATAAAGACCGTCAACCTTATCAATGGCTGGAAATTAATTCAGTT	1628			
Qy	592	GTGGCGCGGTCTACAGCCCGGACTTCTATGAGGACAGGTGCTGTGTGCCAGGCCAAC	651			
Db	1629	GTGGGTGCACTACAGTCCCGAATCTATGAGGACAGGTGCTGTGTGCCAGGCCAAC	1688			
Qy	652	CAGAGGGATCTTCATGTACAGTGGCCCGAGTTCACTGTATATATCTTCTGCAATGCCT	711			
Db	1689	CAGAGGGATCTTCATGTATACAGTGGCCCGAGTTCACTGTATATATCTTCTGCAATGCCT	1748			
Qy	712	GGCTTTCATATCCGCGCGGCACTGCTGAGCTGTGATACCGAGGGGCTCACCTTCGAGGC	771			
Db	1749	GGCTTTCGCTATCCAGAGCCACCGCGCGCGCCCTACCGAGGGGCGCACCTTCGAGGC	1808			
Qy	772	CGTGGTGCACCGGTGTACAACTTTACAGCTGTGGCGCGCCGCCAACCTCCCGGCC	831			
Db	1809	CGCGGTGCACCGGTGTACAACTTTACAGCTGTGGCGCGCGCCGCCAACCTCCCGGCC	1868			
Qy	832	TATGGCGGAGTAGTGATCAAGAGCCAGTGTATGGCAATAAATGCTACAGGGTGGTTAC	891			

gene 1..2279
CDS /gene="A2Bp"
1047..2180 /db_xref="GI:8671586"
/gene="A2Bp"
/codon_start=1
/product="ataxin 2-binding protein"
/protein_id="AAF78291.1"
/db_xref="GI:8671586"
/translation="MAOPYASQAFAPONGIPAEYTAPHHPAPEYTGQTTPVEHTLN
LYPAQTHSESPADTSYAOTVSGTATDQDDPOTGPSTENKSPQKRLHVS
NIPRFRDPDLROMEGOGFKTLIDVELIENRGSKGFVFENSADARAREKLHGTV
VEGSKIENVNATARNTNNKTUNPYTWGKNLNPVVGVAVISPEFYAGTVLLUCQAQSGS
SMYSAPSLSVTISAMPGFPPYAPATAAAAYGAHLRGRGTVYNTFERAAPPPPPIPAYG
GVTVDGYFADIYGGVAARYAQPTPATAAAYSDSYGRVYAADPVYHALAPARTYGV
GANMAFLTDTKDRSHADDVLVSLSLQASIYRGVGNRPAY"
1341..1358 /gene="A2Bp"
/misc_feature /note="Region: RNA binding motif RNP-2"
1452..1472 /gene="A2Bp"
/misc_feature /note="Region: RNA binding motif RNP-1"
BASE COUNT 497 a 712 c 636 g 434 t
ORIGIN

Query Match 67.5%; Score 904.2; DB 9; Length 2279;
Best Local Similarity 86.7%; Pred. No. 1.2e-234;
Matches 1117; Conservative 0; Mismatches 123; Indels 49; Gaps 9;

```

Oy      1  GCGGCGGCCCTGCACAAATGGCTCAG-CATTATGCCTCAGCGCAGTTGCTCCACCACGAG 59
Db      1029 GCCGTGCCCCCTGACACAATGGCTCAGCCTTAGCGTTGCGGCCAGTTGCTCCCGCCGAG 1088

Oy      60  AATGGCATCCT- GCAGAAATACACGGCCCTCATCCCCTATCCCGGCCAGAG- ACACCGGC 117
Db      1089 AACGGTATCCCGCGGAATACAGGCCCTCATCCCACCCCGGCCAGAGTACAGGC 1148

Oy     118  CAGGCCACTGTCGCG- CCACACATTAACATTG- ATCTCTCTACACAGACGCACTCGGAG 175
Db     1149 CAGACCACGGTTCGCGAGCACACATTAACCTGTACCTTCGCGCCAGACGCACCTCCGAG 1208

Oy     176  CAGAGC--- GGTATACAAAGGCGACAG- CGGTCTCCGGCACCGCCACACAGACAGATGAT 231
Db     1209 CAGAGCCCGGCGGACACGAGCGCTCAGACCGCTCTCTGGCACCGCCACACAGACAGATGAC 1268

Oy     232  GCCGCCCGCAGCCGCGCAGCCCGACACACAACTTCTGAAAACACAGAAAAACAAGTCC 291
Db     1269 GCAGCAGCAGCGATGCCAGCCCCACACACAACTTCTGAAAACACAGGAAAACAAGTCT 1328

Oy     292  CAGCCCAAGCGGTCGATGTGTCCAATCCCTTCGGTTCCGGATTCAGACCTCCGA 351
Db     1329 CAGCCCAAGCGGTCGATGTGTCCAATATCCCTTCAGTTCCGGATTCGGACCTCAGA 1388

Oy     352  CAATGTTTGCCCAATTGGTAAATATTAGATCTTGAAATTTATTTATGAGCGGGC 411
Db     1389 CAATGTTTGTCAATTGGTAAATCTTAGATGTTGAAATTTATTTATGAGCGAGGC 1448

Oy     412  TCGAAGGGAATTGGTTTTCGTAACTTTCCGAAAATAGTCGGATCGGCACAGGGCGAGGAG 471
Db     1449 TCAAAGGGAATTGGTTTTCGTAACTTTCCGAAAATAGTGCCGATCGGCACAGGGCGAGGAG 1508

Oy     472  AAATTGACGGTACCGTGTGTAGAGGCGGTAAATCGAGGTTTAATATCGCACAGCACGC 531
Db     1509 AAATTACAGCGCACCGTGTGTAGAGGCGGTAAATCGAGGTTAAATATGCCACAGCACGT 1568

Oy     532  GTGATGACTAATAAAAAGGCGGTGAACCCCTACACCAATGGCTGGAAATTAATTCAGTT 591
Db     1569 GTAATGACAAATAAAAAGACCGTCAACCCCTTATACAAATGGCTGGAAATTTGAATCCAGTT 1628

Oy     592  GTGGGCGCGGTCTACAGCCCGGACTTCTATGACGACACGGTGTGTGTGTCAGGCCAAC 651
Db     1629 GTGGGTGCACTCTACAGTCCCGAATCTATGACGACACGGTCTGTGTGTCAGGCCAAC 1688

Oy     652  CAGAGGGGATCTTCATGTACAGTGCGGCCCGAGTTCACTTGTATATATCTTCTGCAATGCCT 711

```

FEATURES
source

Location/Qualifiers
1..1513
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEMBA1003591"
/tissue_type="whole embryo, mainly head"
/clone_lib="HEMBA1"
/dev_stage="embryo, 10 weeks"
/notes="cloning vector: pME18SFL3"
221..1414
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAA91472.1"
/db_xref="GI:7022046"
/translation="MNCEROLRGNOEAAAADPTMAOPYASAOFAAPPQNGIPAEVTAP
HHHPAPEYTGQITVPPHTLNLYPPAQTHSEQPADTSAGTSGTATDADDAAPDTGQP
QTQPSSENTKSPKRLHVSNIPIFRFDPLDQMFQGRKILDOVELIFNERSKRGFG
VTPENSADADRAEKLIHVEGRKIEVNNTATVMTNKTIVNPFYNGWKLNPVGVAV
YSPFVAGTVLLCOANQEGSSMYSPADSPYFPYPAATAAAAYRGAHLRGRG
RTVYNTFRRAAPPPPIPAYGGVYVQDFYGADIYGGYAAVRYAQPATPAATAAAYSDYG
RVYAADPYHALAPATPYGVGMNNAFAPLTDKTRSHADDVGLVLSLQASIVRGYN
RFAPY"

BASE COUNT 354 a 418 c 395 g 346 t
ORIGIN

Query Match 67.68; Score 905.2; DB 9; Length 1513;
Best Local Similarity 86.74; Pred. No. 6.1e-235; Indels 49; Gaps 9;
Matches 1118; Conservative 0; Mismatches 123;

Qy 1 GCGGCGCCCTGACACAAATGGCTCAG-CITATGCGCTCAGCGCAGTGTGCTCCACCCCGAG 59
Db 263 GCGGCTGCGCTGACACAAATGGCTCAGCTTACGGTTCGGGCCAGTTTGTCTCCCGCGAG 322
Qy 60 AATGCGATCCT-GCAGAATACAGCGCCCTCATCCCATCCCGCCGACAG-ACACCGGC 117
Db 323 AACGGTATCCCGCGGAATACAGCGCCCTCATCCCATCCCGCCGACAGTACACAGGC 382
Qy 118 CAGCCCATGTCCCG-CCACACATTAAACTTG-ATCCTCTCTACACACAGCAGCTCCGGAG 175
Db 383 CAGACACAGGTTCCCGAGCACATTAACCTGTACCTCCCGCCACAGCAGCTCCGAG 442
Qy 176 CAGAGC--GCTTATACAGCGCACAG-CGGTCTCCGGACCGCTCTGGCACCGCCACACAGCAGATGAT 231
Db 443 CAGAGCCCGCGGACAGCGCTCAGACCGTCTCTGGCACCGCCACACAGCAGCAGATGAC 502
Qy 232 GCGGCCCGACCGAGCGCCAGCCCGACACACAACTCTGTAACACACAGAAACAAAGTCC 291
Db 503 GCAGCACCAGCGATGGCCAGCCCGCCAGACAACTCTGTAACACACAGAAACAAAGTCT 562
Qy 292 CAGCCCAAGCGCTGCATGTCTCAACATCCCTTCCGGTTCGGGATCCAGACCTCCGA 351
Db 563 CAGCCCAAGCGCTGCATGTCTCAATATCCCTTCCAGTTCGGGATCCGACCTCAGA 622
Qy 352 CAAATGTTTGGCAATTTGGTAAATATATAGATGTTGAAATATTTTAAATGAGCGGGC 411
Db 623 CAAATGTTTGGTCAATTTGGTAAATCTTAGATGTTGAAATATTTTAAATGAGCGAGC 682
Qy 412 TCAGAGGATTTGGTTTGGTAACTTTGGAATAAGTAGTGGGATCGGACAGGCGAGGAG 471
Db 683 TCAAGAGGATTTGGTTTGGTAACTTTGGAATAAGTAGTGGGATCGGACAGGCGAGGAG 742
Qy 472 AAATTCAGCGTACCGTGGTAGAGGCGCGTAAATCGAGTTAATATGCGACGACGACGC 531
Db 743 AAATTCAGCGTACCGTGGTAGAGGCGCGTAAATCGAGTTAATATGCGACGACGACGT 802
Qy 532 GTGATGACTAATAAAGGCGGTGAACCCCTACACCAATGGCTGGAAATTAATATTCAGTT 591
Db 803 GTAATGACAAATAAAGACCGTCAACCCCTTATACAAATGGCTGGAAATTAATATTCAGTT 862
Qy 592 GTGGGCGGTCTACAGCCCGGACTCTATGACGAGCAGCGTGTGTTGTCAGGCGCAAC 651
Db 863 GTGGGCGGTCTACAGTCCCGAAATCTATGACGAGCAGCGTCTGTTGTCAGGCGCAAC 922

Qy 652 CAGGAGGATCTTCATGTACAGTGGCCCGAGTTCACCTGTATATACCTTCTGCAATGCT 711
Db 923 CAGGAGGATCTTCATGTACAGTGGCCCGAGTTCACCTGTATATATCTTCTGCAATGCA 982
Qy 712 GCGTTTCATATCGGCGCCACCTGCTGCAGCTGCATACCGAGGGGCTCACCTTCGAGGC 771
Db 983 GCGTTTCGATTCAGAGAGCCACCGCGGGCGGCTTACCGAGGGGCGACCTCGGAGGC 1042
Qy 772 CCGTGTCCGACCGCTTACAACACCTTTCAGAGCTGCGGGCGGCCACCCCACTCCCGGCC 831
Db 1043 CCGGTTCCGACCGCTTACAACACCTTTCAGGCGCGGGCGGCCCGCCCGGCTCCCGGCC 1102
Qy 832 TATGCGGAGTAGTGTATCAAGAGCCAGTGTATGCAATAAATTCGTACAGGGTGTGTAC 891
Db 1103 TACGGCGGTGTTGTATACAGGATGGAATTTATGG---TGCAGACATTTATGTTGTTAT 1159
Qy 892 GCTGCATACCGCTAGCGCCAGCCACCCCTGCACTGCTGTGCTTACAGTACAGTACAGTTAC 951
Db 1160 GCTGCATACCGCTAGCGCCAGCTTACCCCTGCACTGCGGCTGCCTACAGTACAGTTAC 1219
Qy 952 GGAAGAGTTTATGCTGCCGACCCCTTACCAACACACTTGTCTCCAGCCCGCCACCTACGGC 1011
Db 1220 GGAAGAGTTTATGCTGCCGACCCCTTACCAACACACTTGTCTCCAGCCCGCCACCTACGGC 1279
Qy 1012 GTTGTGCCATGATGCTTGTGCGCCCTTACCGATGCGCAAGACTAGGAGCCATGCTGAT 1071
Db 1280 GTTGTGCCATGATGCTTGTGCGCCCTTACCGATGCGCAAGACTAGGAGCCATGCTGAT 1339
Qy 1072 GATGTGGTCTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1131
Db 1340 GATGTGGTCTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1399
Qy 1132 TTTGCTCCATATTAATGATAAAACCATTAACAAACAGCAAAACAAACAAACAAACAA 1191
Db 1400 TTTGCTCCATATTAATGATAAAACCATTAACAAACAGCAAAACAAACAAACAAACAA 1430
Qy 1192 AAAAACCAACCTTCCAACTGTGGGAGAGAGAAAGCTTTCGAGGCGCCAGTGTTCGCAC 1251
Db 1431 -----AACTTCCATGTGGGAGAGAGAAAGCTTTCGAGGCGCTGAGTATTCGAAT 1482
Qy 1252 ACATGAGTAGACATCATTATAGCAACTC 1281
Db 1483 ACATGAGTAGACATCATTATAGCAACTC 1512

RESULT 8
AF107203 2279 bp mRNA linear PRI 23-JUN-2000
LOCUS Homo sapiens ataxin 2-binding protein (A2BP) mRNA, complete cds.
DEFINITION
ACCESSION AF107203
VERSION AF107203.1 GI:8671585
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Shibata,H., Huynh,D.P., Vo,T.T. and Pulst,S.-M.
TITLE A novel protein, A2BP, with RNA binding motif binds to C-terminal ataxin-2
JOURNAL Unpublished
REFERENCE
AUTHORS Shibata,H., Huynh,D.P., Vo,T.T. and Pulst,S.-M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1998) Division of Neurology and Neurogenetics Laboratory, Burns and Allen Research Institute, Cedars-Sinai Medical Center, UCLA School of Medicine, 8700 Beverly Blvd., Los Angeles, CA 90048, USA
FEATURES
1..2279
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="similar to Homo sapiens cDNA clone HSP38A20"

/translation="MLASQGVLLHPYGVPMIVPAAPYLPLGIQNGEAAAPDITMAQF
YASQAPQNGIPAEYTAHPHPAPEYTGOTTPERTLNLYPAOTHSOSPADTSA
QVSGTATOTDDAAPTQOPQPSSENTENKSPQLRHVSNIPFRPDLPDRMFQGF
GKILDEIIFNERSKGFVTFENSADARERKJHCTVVEGRKIEVNNATARVTN
KKTVPYTNKWLNPVVGAVSPFYAGTVLLCOANGEGSMYSAPSLYVTSMPGF
PVPATAAAVGAHLRGRGTVYNTFRMAAPPPIFAYGSVVYQEBYGNKLLQGGY
AAYRAYQPTPATAAAYSDRNFVFAADEISNTSAVTDFMLPTPTTTHLOPPPTA
LVP"
polyA_signal 3299...3304
/gene="HRNP1"
BASE COUNT 950 a 775 c 711 g 912 t
ORIGIN

Query Match 67.6%; Score 906.4; DB 9; Length 3348;
Best Local Similarity 84.4%; Pred. No. 3.3e-235;
Matches 1165; Conservative 0; Mismatches 116; Indels 99; Gaps 9;

QY 1 GCGGCGCCCTTGACACAAATGGCTCAG-CTTATGCTCAGCGCAGTTCGCTCCACCCAG 59
DB 356 GCCGTGCCCCTGACACAAATGGCTCAGCGCTTCGCGCCAGTTTGTCTCCCGCCGAG 415
QY 60 AATGCGATCCT-GCAGAATACAGCGCCCTCATCCCATCCCGGCCAGAG-ACACCGGC 117
DB 416 AACGGTATCCCGGGAATACAGCGCCCTCATCCCAACCCCGGCCAGAGTACACAGGC 475
QY 118 CAGCCCATGTCCCGG-CCACACATTAACCTTG-ATCCTCTCACAGACGCACTCGGAG 175
DB 476 CAGACCAAGGTTCCCGAGCACATTAACCTGTACCTCTCCCGCCAGAGCGACTCCGAG 535
QY 176 CAGAGC---GCTTATACAGCGGCAG-CCGTCTCCCGGACCGGCACACAGACGATGAT 231
DB 536 CAGAGCCCGGCGACAGCGCTCAGACCGCTCTCGGACCGCCACACAGACGATGAC 595
QY 232 GCCGCCCGCAGCGGCCCGCCAGACCAACCTTCTGAAACACAGAAACAACTCC 291
DB 596 GCAGCAGCGGATGCGCAGCCCGCCAGACCAACCTTCTGAAACCGGAAACAGTCT 655
QY 292 CAGCCCAAGCGCTGCATGTGTCAACATCCCTTCGGTTCGGGATCCAGACCTCCGA 351
DB 656 CAGCCCAAGCGGTCATGTGTCCAAATATCCCTTCAGGTTCCGGATCCGGACCTCAGA 715
QY 352 CAAATGTTTGCCCAATTTGGTAAATATTAGATGTTCAAAATTTTATATAGCGGGC 411
DB 716 CAAATGTTTGTCATTTGGTAAATTTAGATGTTGAAATTTTATATAGCGAGGC 775
QY 412 TCGAGGATTTGTTTCGTAATCTTCGAAAATAGTCGGATCGGACGAGCGGAG 471
DB 776 TCAAAGGATTTGGTTTCGTAATCTTCGAAAATAGTCGGATCGGACGAGCGGAG 835
QY 472 AAATTGACGATACCGGTAGAGGCGGTAAATTCGAGTTAAATTCGGACAGCACGC 531
DB 836 AAATTACCGGACCGGTGGTAGAGGCGGTAAATTCGAGTTAAATTCGACAGCACGT 895
QY 532 GTGATGACTAATAAAGGCGGTGAACCCCTACCAATGCGTGGAAATTAATCCAGTT 591
DB 896 GTAATGACAAATAAAGACCGGTCAACCTTATACAAATGCGTGGAAATTAATCCAGTT 955
QY 592 GTGGGCGGTGTACAGCCCCGACTTCTATGACGACCGGTGCTGTGTGCCAGGCCAAC 651
DB 956 GTGGGTGCGATGTACAGTCCGAAATCTATGACGACCGGTCTGTGTGCCAGGCCAAC 1015
QY 652 CAGGAGGATCTTCCATGTACAGTGGCCCGCAGTTCACTTGTATATCTTCTCGAATGCT 711
DB 1016 CAGGAGGATCTTCCATGTACAGTGGCCCGCAGTTCACTTGTATATCTTCTCGAATGCCA 1075
QY 712 GCGTTTCCATATCCGCGCCCACTGCTGCGAGTGTGATACCGAGGGCTCACCTTCGAGGC 771
DB 1076 GCGTTCCGATCCAGCGCACCGCGCGCGCTACCGAGGGGCGCACCTTCGAGGC 1135
QY 772 CGTGGTCCGACGTTGACACACTTTCAGAGTTCGGGCGCCCGCCCAATCCCGGCC 831
DB 1136 CCGGCTCGACCGTGTACAAACACTTTCAGGCGCGGGCGCGCCCGCCCGATCCCGGCC 1195

QY 832 TATGCGGAGTAGTGTATCAAGACCGAGTGTATGGCAATAAATGCTTACAGGTTGTTAC 891
DB 1196 TACGCGGAGTAGTGTATCAAGACCGCTGTGTATGGCAATAAATGCTTACAGGTTGTTAT 1255
QY 892 GCTGCATACCGCTACGCGCCACCGCCCTGCGCACTGCTGCTGCTTACAGTGA----- 944
DB 1256 GCTGCATACCGCTACGCGCCACCGCTACCCCTGCGCACTGCGCGCTGCTTACAGTGA 1315
QY 945 -----CAGTTACGCGACGAG 958
DB 1316 CAGTTGCTCTTCTGTCAGCAGATGAATAATTTCTTAAACACCTCTGTCAGTTACGCGAG 1375
QY 959 TTTATGCTGCGGACCGCTTACCCACACACACTTGTCTCCAGCCCGCCACTTACGCGGTTGGTG 1018
DB 1376 TTTATGCTGCGGACCGCTTACCCACACACTTGTCTCCAGCCCGCCACTTACGCGGTTGGTG 1435
QY 1019 CCATGAATGCTTTTGGCGCCCTTGACCGCATGCCAAGACTAGGAGCCCATGCTGATGATGTGG 1078
DB 1436 CCATGAATGCTTTTGGCACTTTGACTGATGCCAAGACTAGGAGCCCATGCTGATGATGTGG 1495
QY 1079 GTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1138
DB 1496 GTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1555
QY 1139 CATATTAAATGATAAAACCATTAACCAACAAACAAACAAACAAACAAACAAACAAAC 1198
DB 1556 CATACTAAATGACAAACCATTAAC----- 1579
QY 1199 CAACTTCCAAATGTGGGAGAGAGGAAGCTTTCCGAGGCCCGAGTGTTCGACACATGCA 1258
DB 1580 -AACCTTCCAAATGTGGGAGAGGAAGCTTTCCGAGGCCCGAGTGTTCGACATATGCA 1638
QY 1259 GTAGGACATCACTTTTACCACTCAAGAAACAAACAAACAAACAAACAAACAAACAA 1318
DB 1639 GTAGTACATCATTTTACCACTCTTAAACAAACAAACAAACAAACAAACAAACAAAC 1698

RESULT 7
AK001027

LOCUS 1513 bp mRNA linear PRI 01-AUG-2002
DEFINITION Homo sapiens cDNA FLJ10165 fis, clone HEMBA1003591, weakly similar
to CHLOROPLAST 28 KD RIBONUCLEOPROTEIN PRECURSOR.
ACCESSION AK001027
VERSION AK001027.1 GI:7022045
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens embryo, 10 weeks whole embryo, mainly head cDNA to
mRNA, clone lib:HEMBA1 clone:HEMBA1003591.

ORGANISM

Homo sapiens
Eukaryote; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H.,
Wagatsuma, M., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K.,
Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K.,
Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A.,
Nakamura, Y., Nagahara, K., Maeho, Y., Ninomiya, K. and Iwayanagi, T.
NEDO human cDNA sequencing project

REFERENCE

1
Isogai, T. and Otsuki, T.
Direct Submission
Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection;
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.

TITLE

Unpublished
2 (bases 1 to 1513)
Isogai, T. and Otsuki, T.
Direct Submission

JOURNAL

Isogai, T. and Otsuki, T.
Direct Submission

AUTHORS

Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H.,
Wagatsuma, M., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H.,
Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K.,
Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K.,
Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A.,
Nakamura, Y., Nagahara, K., Maeho, Y., Ninomiya, K. and Iwayanagi, T.
NEDO human cDNA sequencing project

COMMENT


```
LYPTOTHSQSADTSAGTSGTATGDDAAPTGGQPTQPSNTESKSPKRLHVSN
IPFRFRDPLRQMGFGKILDEIIFNERSGKGFVTFENSADADRARKLHGT
VBERKLENNATATVNTVNPVKNLNPVGVAVSPDFYAGTVLLCOANOBG
SSMTYSGSSLYTTSAMPGFPYPAATATAAAVGAHLRGRTRVNTFRAPAAPBPPIPAY
GGVYQDGFYGADLYGGVAAVYAAQPTPATAAAVSDSYGRVYAADPVHHTLAPPTVG
VGMNAPAPLTDKTRSHADDVGLVLSLQNSIYRGYNRFAPY"
misc_feature      358..375
/gene="A2bp"
/note="Region: RNA binding motif RNP-2"
475..495
/gene="A2bp"
/note="Region: RNA binding motif RNP-1"
BASE COUNT      362 a 396 c 324 g 281 t
ORIGIN
Query Match      80.9%; Score 1084.4; DB 10; Length 1363;
Best Local Similarity 93.6%; Pred. No. 1e-283;
Matches 1233; Conservative 0; Mismatches 66; Indels 19; Gaps 9;
QY 1 GCGCGCGCCCTGACACAAATGGCTCAG-CTTATGCTCAGCGCAGTTGCTCCACCCCGAG 59
DB 49 GCGCGCGCCCTGACACAAATGGCTCAGCTTATGCTCAGCGCAGTTGCTCCACCCCGAG 108
QY 60 AATGCGCAT-CTGCGAATAACACGCGCCCTCATCCCATCCGCGCCGAGAG-ACACCGCG 117
DB 109 AATGCGCATCTGAGAAATACACGCGCCCTCATCTCTATCCGCGCCGAGAGTACACCGCG 168
QY 118 CAGCCCACTGTCCCG-CCACACATTAACTTG-ATCCTCTCACACAGCGCACTCGGAG 175
DB 169 CAGACCACTGTCCCTGACACACATTAAACCTGTATCTCTCTACACAGCGCACTCGGAG 228
QY 176 CAGAGCCCTTATACAAAGGACAG-CGCTCTCGGACCGCCACACAGACAGATGATGCC 234
DB 229 CAGAGTGTGACACACAGTGGCGACACCGTCTCCGCGCACCGCCACACACAGATGATGCA 288
QY 235 GCCCGGACCGACGCGCCACACACACACACCTTCTGAAACACAGAAACAGTCCCGAG 294
DB 289 GCCCGGACCGAGTGGCGAGCCCGACACACACCTTCTGAAACACAGAAACAGTCCCGAG 348
QY 295 CCAAGCGGTGCATGTGTCAACATCCC-----CTTCGCGTTCCGGGATCCGACGCTC 348
DB 349 CCAAGCGGTGCATGTGTCAACATCCCCTTTCGCGTTCCGGGATCCGACGCTC 408
QY 349 CGACAAATGTTGGCCCAATTTGGTAAATATTAGATGTTGAAATTTATTAATAGCGG 408
DB 409 CGACAAATGTTGGTCAATTTGGTAAATATTAGATGTTGAAATTTATTAATAGCGCA 468
QY 409 GGCTCGAAGGGATTGGTTTCGTAACCTTCGAAAAATAGTCGCGATCGGACGCGGAGG 468
DB 469 GGCTCGAAGGGATTGGTTTCGTAACCTTCGAAAAATAGTCGCGATCGGACGCGGAGG 528
QY 469 GAGAAATTGACGGTACCGTGTAGAGGCGCGTAAATCGAGTTAATATGCGACAGCA 528
DB 529 GAGAAATTGACGGTACCGTGTAGAGGCGCGTAAATCGAGTTAATATGCGACAGCA 588
QY 529 CGCGTGATGACTAATAAAGGCGCGTGAACCCCTCACCAATGGCTGGAATTAATCCA 588
DB 589 CGCGTGATGACTAATAAAGGCGCGTGAACCCCTCACCAATGGCTGGAATTAATCCA 648
QY 589 GTTGTGGCGCGGTCTACAGCCCCGACTTCTATGACGAGCACGCTGCTGTGTGCGCAGGC 648
DB 649 GTTGTGGCGCGGTCTACAGCCCCGACTTCTATGACGAGCACGCTGCTGTGTGCGCAGGC 708
QY 649 AACGAGGAGGATCTTCATGTACAGTGGCCCGAGTTCACTTGTATATCTTCTGCAATG 708
DB 709 AACGAGGAGGATCTTCATGTACAGTGGCCCGAGTTCACTTGTATATCTTCTGCAATG 768
QY 709 CTGTGCTTTCCATATCCGCGCCGCTGCTGAGCTGCATACCGAGGGGCTCACCTTGA 768
DB 769 CTGTGCTTTCCATATCCGCGCCGCTGCTGAGCTGCATACCGAGGGGCTCACCTTGA 828
QY 769 GCGCGGTGTCGACCGTGTACACACCTTTCAGAGTTCGCGCGCCCGCCACCCCAATCCG 828
```

```
DB 829 GCGCGTGGTCCGACCGTGTAACACACCTTCAGGGCTGCAGCGCCCGCCCAATCCCG 888
QY 829 GCTATGGCGGAGTACTGTATCAAGACCGAGTGTATGGCAATAAATTTGTACAGGGTGGT 888
DB 889 GCTATGGCGGAGTGTGTATACCGAGGATGGAATTTATGG---TGACAGACATTTATGGTGGT 945
QY 889 TAGCGTGCATACCGCTACGCGCCGACCGCCCTGCGCACTGCTGCTCCTACAGTGACAGT 948
DB 946 TATGCTGGCTACCGCTATGCGCAGCGCCACCCCTGCGCACTGCGCTGCCCTACAGTGACAGT 1005
QY 949 TAGCGACGAGTTTATGCTGCGGACCCCTTACCACACACACTTGTCTCAGCCCCCAGCTTAC 1008
DB 1006 TAGCGACGAGTTTATGCTGCGGACCCCTTACCACACACACTTGTCTCAGCCCCCAGCTTAC 1065
QY 1009 GCGTGTGGTGGCCATGATGCTTTTGGCGCCCTTGACCGATGCCAAGCTAGGAGCCATGCT 1068
DB 1066 GCGTGTGGTGGCCATGATGCTTTTGGCGCCCTTGACCGATGCCAAGCTAGGAGCCATGCT 1125
QY 1069 GATGATGTGGTCTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1128
DB 1126 GATGATGTGGTCTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1185
QY 1129 CGTTTGTGCTCCATTTAAATGATAAAACCATTTAAACAAACAAACAAACAAACAAACAAAC 1185
DB 1186 CGTTTGTGCTCCATTTAAATGATAAAACCATTTAAACAAACAAACAAACAAACAAACAAAC 1245
QY 1186 -AAAAACAAACAAACCAACCTTCCCAATGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1244
DB 1246 GAAACAAACAAACCAACCTTCCCAATGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1305
QY 1245 TTGCGACACATGACGATAGGACATCACTTTAGCAACTCAAAAGAAACAAACGAAAAA 1302
DB 1306 TTGCGACATGACGATAGGACATCACTTTAGCAACTCAAAAGAAACAAACGAAAAA 1363
```

```
RESULT 5
AF229057 Homo sapiens hexaribonucleotide binding protein 1 isoform gamma
LOCUS AF229057 2000 bp mRNA linear PRI 01-FEB-2002
DEFINITION Homo sapiens hexaribonucleotide binding protein 1 isoform gamma
ACCESSION AF229057 (HRNBPI) mRNA, complete cds.
VERSION AF229057.1 GI:18461366
KEYWORDS SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Chen, W. and Winkelman, J.C.
Submitted (28-JAN-2000) Internal Medicine/Hematology-Oncology, for
University of Cincinnati College of Medicine, The Vontz Center for
Molecular Studies, 1215 Eden Avenue, Cincinnati, OH 45267-0508, USA
Location/Qualifiers
1. 2000
/organism="Homo sapiens"
/db xref="taxon.9606"
/chromosome="16"
/map="16p13.3"
1. 2000
/gene="HRNBPI"
256..1512
/notes="RNA-binding protein"
/codon_start=1
/product="hexaribonucleotide binding protein 1 isoform
gamma"
protein_id="AAL71904.1"
/db xref="GI:18461367"
/translaton="MLASGVLHPVPMIVPAAAPLPLGLICGNQEAAPDTPMAQP
YASAPAPONGIPAEYTAHPHAPPEYTGQTTVPETHLNLIPPAQTHSEQPADTSA
QTVSGATQTDNAAPTDDGQPTQPSNTENKSPKRLHVSNIIPFRDPLRQMGFGK
GILDEIIFNERSGKGFVTFENSADADRARKLHGTVVEGRKIEVNNATRVMTN
```

source

gene

CDS


```
/dev_stage="adult"
637...1827
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAA95079.1"
/db_xref="GI:7670456"
/translation="MNCEREQLRGNQEAAPDTPAQPVASQAQFAPPONGIPAEVTAP
HPHPEYTGQTPVGHNLVPTTQTHSEQADTSQATVSGTATQDDAQTDSQPO
TOPSENTESKQPKRLHVSNIIPRRPDLROMFGQFQKLLDVELIPNERSKRGFV
TPENSADADAREKLGHTVEGRKLEVNATARVWINKTVNPTNGWKLNPVGVAVY
SPDYACTVLLCOANQEGSSMYGSSSLVTSAMPGFPYPAATAAAVYGAHLRGRGR
TYVNTFRRAAPPPIPAYGVVYQDFYAGDIYGGYAAVRYAQPTPATAAAYSISYGR
VYAADPYHHTLAPPTYGVGMNAPFLTDKTRGHADDVGLVLSLQASIVRGYNR
FAPY"
BASE COUNT      518 a   589 c   516 g   379 t
ORIGIN

Query Match      82.8%; Score 1110; DB 10; Length 2002;
Best Local Similarity 94.1%; Pred. No. 1.2e-290;
Matches 1242; Conservative 0; Mismatches 65; Indels 13; Gaps 8;

QY 1 GCGGCGCCCTGACACATGGCTCAG-CATTATGCTTCAGCGAGTTCGCTCCACCCCGAG 59
DB 679 GCGGCGCCCTGACACATGGCTCAGCTTATGCTTCAGCGAGTTCGACACCCCGAG 738
QY 60 AATGGCAT-CTTCAGAAATACAGCGCCCTCATCCCATCCCGCGCCAGAG-ACACCGCG 117
DB 739 AATGGCATCCCTGACAAATACAGCGCCCTCATCCCTCATCCCGCGCCAGTACACCGCG 798
QY 118 CAGCCCACTGTCCC-CCGCACACATTAACCTTG-ATCCCTCCTACACAGACGACTCGGAG 175
DB 799 CAGACCACTGTCCCTGGCCACACATTAACCTGTATCCCTTACACAGACGACTCGGAG 858
QY 176 CAGAGCGCTTATACAGCGCCAG-CGCTCTCCGCGACCGCCACACACAGACATGATGCC 234
DB 859 CAGAGTGTCTGACACCATGCGCAGACCGTCTCCGCGACCGCCACACAGACATGATGCA 918
QY 235 GCGCCGACCGCGCCAGCCGACACAGACAACTTCTGAACACAGAGAAACAAAGTCCCGAG 294
DB 919 GCGCCGACCGCGCCAGCCGACACAGACAACTTCTGAACACAGAGAAACAAAGTCCCGAG 978
QY 295 CCCAAGCGCTGCATGTGTCCAAACATCCCTTCCGGTTCGGGATCCAGACTCCGACAA 354
DB 979 CCCAAGCGCTGCATGTGTCCAAACATCCCTTCCGGTTCGGGATCCAGACTCCGACAA 1038
QY 355 ATCTTTGGCCAAATTTGGTAAATATATAGATGTGAAATATATTTTAAATGAGCGGGCTCG 414
DB 1039 ATGTTTGGTCAATTTGGTAAATATATAGATGTGAAATATATTTTAAATGAGCGGGCTCC 1098
QY 415 AAGGGATTTGGTTTCGTAACTTTCGAAATAGTGGGATCCGAGCGGAGCGGAGAGAA 474
DB 1099 AAGGGATTTGGTTTCGTAACTTTCGAAATAGTGGGATCCGAGCGGAGCGGAGAGAA 1158
QY 475 TTGCAACGTTACCGTGTAGAGCGCGTAAATTCGAGGTAAATTCGACAGACGCGGTG 534
DB 1159 TTGCAACGTTACCGTGTAGAGCGCGTAAATTCGAGGTAAATTCGACAGACGCGGTG 1218
QY 535 ATGACTAATAAAGCGCGTGAACCCCTACACCAATGCTCGAAATTAATTCAGTTGTG 594
DB 1219 ATGACAAATAGAGACTGTCAACCCCTTACACCAATGCTCGAAATTAATTCAGTTGTG 1278
QY 595 GCGCGGTCTACAGCCCGACTTCTATCAGCGACGCGTGTGTTGTCGAGCGCAACCGAG 654
DB 1279 GCGCGGTCTACAGCCCGACTTCTATCAGCGACGCGTGTGTTGTCGAGCGCAACCGAG 1338
QY 655 GAGGATCTTCATGTACAGTGGCCCGAGTTCACTTGTATATATCTTCTGCAATGCTGGC 714
DB 1339 GAGGATCTTCATGTACAGTGGCCCGAGTTCACTTGTATATATCTTCTGCAATGCTGGC 1398
QY 715 TTTCCATATCCGGCGGCACCTGCTGACGCTCATACCGAGGGGCTACCTTCGAGGCGGT 774
DB 1399 TTTCCATATCCGGCGGCACCTGCTGACGCTCATACCGAGGGGCTACCTTCGAGGCGGT 1458
```

```
QY 775 GGTCCACCGTGTACACACACCTTCAGAGCTCGCGGCGCCACCCCAATCCCGGCTTAT 834
DB 1459 GGTCCACCGTGTACACACACCTTCAGAGCTCGCGGCGCCACCCCAATCCCGGCTTAT 1518
QY 835 GCGGAGTAGTGTATCAAGAGCCAGTGTATGCAATAAATGCTACAGGTTGGTTAGCT 894
DB 1519 GCGGAGTAGTGTATCAAGAGTGGATTTTATGG---TGACAGACATTTATGGTTTATGCT 1575
QY 895 GCATACCGCTACGCCACGCCACCCCTGCCACTGCTGCTGCTACAGTGACAGTTACGGA 954
DB 1576 GCGTACCGCTATGCCAGCCACCCCTGCCACTGCGCTGCTACAGTGACAGTTACGGA 1635
QY 955 CGAGTTATGCTGCGGACCCCTACACACACACTTCTGCTCAGCCCCCACTACGGCGTT 1014
DB 1636 CGAGTTATGCTGCGGACCCCTACACACACACTTCTGCTCAGCCCCCACTACGGCGTT 1695
QY 1015 GGTGCATGAATGCTTTTGGCGCCCTTGACCGATGCCAAGACTAGGAGCCCATGCTGATGAT 1074
DB 1696 GGTGCATGAATGCTTTTGGCGCCCTTGACCGATGCCAAGACTAGGAGCCCATGCTGATGAT 1755
QY 1075 GTGGTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1134
DB 1756 GTGGTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1815
QY 1135 GCTCCATATTAAATGATAAAACCAATTAACAAACAGCAAAACAAACAAACAAACAAACAA 1190
DB 1816 GCTCCATATTAAATGATAAAACCAATTAACAAACAGCAAAACAAACAAACAAACAAACAA 1875
QY 1191 AAAAAACCAACCTTCAATGTGGGAGAGAGGAAGCTTTCCGAGGCCCGAGTGTTCGGA 1250
DB 1876 AAAAAACCAACCTTCAATGTGGGAGAGAGGAAGCTTTCCGAGGCCCGAGTGTTCGGA 1935
QY 1251 CACATGCACTAGACATCACTTTAGCACTCAAGAAACAAACGAAAAAAGAAAAA 1310
DB 1936 CACATGCACTAGACATCACTTTAGCACTCAAGAAACAAACGAAAAAAGAAAAA 1995
```

```
RESULT 4
AF107204 LOCUS 1363 bp mRNA linear ROD 23-JUN-2000
DEFINITION Mus musculus ataxin 2-binding protein (A2bp) mRNA, complete cds.
ACCESSION AF107204
VERSION AF107204.1 GI:8671587
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Shibata,H., Huynh,D.P., Vo,T.T. and Pulst,S.-M.
TITLE A novel protein, A2BP, with RNA binding motif, binds to C-terminal
ataxin-2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1363)
AUTHORS Shibata,H., Huynh,D.P., Vo,T.T. and Pulst,S.-M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1998) Division of Neurology and Neurogenetics
Laboratory, Burns and Allen Research Institute, Cedars-Sinai
Medical Center, UCLA School of Medicine, 8700 Beverly Blvd., Los
Angeles, CA 90048, USA
FEATURES
source
1..1363
/organism="Mus musculus"
/db_xref="taxon:10090"
gene
1..1363
/gene="A2bp"
CDS
67..1203
/gene="A2bp"
/codon_start=1
/product="ataxin 2-binding protein"
/protein_id="AAF78292.1"
/db_xref="GI:8671588"
/translation="MAQPYASAQFAPPQNGIPAEYTAHPHPHPEYTGQTTVPDHTLN
```

```
/clone lib="Sugano mouse brain mncb"
```


JOURNAL Patent: WO 0174901-A 1 11-OCT-2001;

Scios Inc. (US)

FEATURES Location/Qualifiers

source 1..1340

/organism="Rattus norvegicus"

/db_xref="taxon:10116"

BASE COUNT 366 a 394 c 317 g 263 t

ORIGIN

Query Match 100.0%; Score: 1340; DB 6; Length 1340;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGGCGCCCTGACAAATGGCTCAGCTTATGCCTCAGCGCAGTTTCGCTCCACCCCGA 60
DB 1 GCGGCGCCCTGACAAATGGCTCAGCTTATGCCTCAGCGCAGTTTCGCTCCACCCCGA 60
QY 61 ATGGCATCTGCAAGATACACGCGCCCTCATCCCATCCCGCCGACAGACACCGGCCAG 120
DB 61 ATGGCATCTGCAAGATACACGCGCCCTCATCCCATCCCGCCGACAGACACCGGCCAG 120
QY 121 CCACATGTCCTCCGCCACACATTAACCTTGATCTCTCTACACAGACGCACTCGGAGCAGAG 180
DB 121 CCACATGTCCTCCGCCACACATTAACCTTGATCTCTCTACACAGACGCACTCGGAGCAGAG 180
QY 181 CGCTTATACAGCGCAGCGCTCTCCGGACCGCCACACAGACAGATGATCCGCCCGG 240
DB 181 CGCTTATACAGCGCAGCGCTCTCCGGACCGCCACACAGACAGATGATCCGCCCGG 240
QY 241 ACCGACGCGCCGACGACACAACTCTCTGAAACACAGAAACAAAGTCCCGACCCCAAG 300
DB 241 ACCGACGCGCCGACGACACAACTCTCTGAAACACAGAAACAAAGTCCCGACCCCAAG 300
QY 301 CGCTGCATGTGTCACATCCCTTCCGGTTCGGGATCCAGACCTCCGACAAATGTTT 360
DB 301 CGCTGCATGTGTCACATCCCTTCCGGTTCGGGATCCAGACCTCCGACAAATGTTT 360
QY 361 GCGCAATTTGGTAAATATAGATCTGAAATATTTTAAATGAGCGGGCTCGAAGGGA 420
DB 361 GCGCAATTTGGTAAATATAGATCTGAAATATTTTAAATGAGCGGGCTCGAAGGGA 420
QY 421 TTTGGTTTCGTAACTTTCGAAATAGTCGGATCGGACGAGCGGAGGAGAAATTCAC 480
DB 421 TTTGGTTTCGTAACTTTCGAAATAGTCGGATCGGACGAGCGGAGGAGAAATTCAC 480
QY 481 GGTACCGTGTAGAGGCGGTAATATCGAGGTTAATAATGCGACAGCAGCGTGATGACT 540
DB 481 GGTACCGTGTAGAGGCGGTAATATCGAGGTTAATAATGCGACAGCAGCGTGATGACT 540
QY 541 AATAAAGGCGGTGACCCCTACACCAATGCTGGAATTAATCCAGTTGTGGGCGG 600
DB 541 AATAAAGGCGGTGACCCCTACACCAATGCTGGAATTAATCCAGTTGTGGGCGG 600
QY 601 GTCTACAGCCCGACTTCTATGAGGCGAGTGTGTTGTCGAGGCCAACACGAGGGA 660
DB 601 GTCTACAGCCCGACTTCTATGAGGCGAGTGTGTTGTCGAGGCCAACACGAGGGA 660
QY 661 TCTTCCATGTACAGTGGGCCCGGCTACCTTGTATATCTCTGCAATCCCTGGCTTCCA 720
DB 661 TCTTCCATGTACAGTGGGCCCGGCTACCTTGTATATCTCTGCAATCCCTGGCTTCCA 720
QY 721 TATCCGGCGCCGCTGCTGACGCTGCATACCGAGGGGCTCACCTTCGAGGCGGTGTCG 780
DB 721 TATCCGGCGCCGCTGCTGACGCTGCATACCGAGGGGCTCACCTTCGAGGCGGTGTCG 780
QY 781 ACCGTGTACACACTTTCAGAGCTGCGGCGCCCGACCCCAATCCCGGCTATGGCGGA 840
DB 781 ACCGTGTACACACTTTCAGAGCTGCGGCGCCCGACCCCAATCCCGGCTATGGCGGA 840
QY 841 GTAGTGTATCAAGACGAGTGTAGGCAATTAATGCTACAGGCTGTTTACGCTGCATAC 900
DB 841 GTAGTGTATCAAGACGAGTGTAGGCAATTAATGCTACAGGCTGTTTACGCTGCATAC 900

QY 901 CGCTAGCCGCA3CCACCCCTGCCACTGCTGCTGCTACAGTACAGTACGAGGAGTT 960
DB 901 CGCTAGCCGCA3CCACCCCTGCCACTGCTGCTGCTACAGTACAGTACGAGGAGTT 960
QY 961 TATGCTGCGGACCCCTACCAACACACACTTCTCAGGCCCCCCTACGCGGTTGGTGCC 1020
DB 961 TATGCTGCGGACCCCTACCAACACACACTTCTCAGGCCCCCCTACGCGGTTGGTGCC 1020
QY 1021 ATGAATGCTTTTGGCCCTTACCGATGCGCAAGCTAGGAGCCATGCTGATGATGGT 1080
DB 1021 ATGAATGCTTTTGGCCCTTACCGATGCGCAAGCTAGGAGCCATGCTGATGATGGT 1080
QY 1081 CTCGTTCTTCTTCTCATTTGAGGCTAGTATATACCAAGGGGATACAAACGTTTGGTCCA 1140
DB 1081 CTCGTTCTTCTTCTCATTTGAGGCTAGTATATACCAAGGGGATACAAACGTTTGGTCCA 1140
QY 1141 TATTAATGATATAAACCATTAAACAAACAGCAAAACAAACAAACAAACAAACAAACCA 1200
DB 1141 TATTAATGATATAAACCATTAAACAAACAGCAAAACAAACAAACAAACAAACAAACCA 1200
QY 1201 ACCTTCAATGTCGGGAGAGGAGGAGCTTCCGAGGCCGAGTGTTCGGACACATGAGT 1260
DB 1201 ACCTTCAATGTCGGGAGAGGAGGAGCTTCCGAGGCCGAGTGTTCGGACACATGAGT 1260
QY 1261 AGGACATCCTTTAGCAACTCAAGAAACAAACAAACAAACAAACAAACAAACAAACCA 1320
DB 1261 AGGACATCCTTTAGCAACTCAAGAAACAAACAAACAAACAAACAAACAAACAAACCA 1320
QY 1321 GCGCGAAGGGTTCGCTAGA 1340
DB 1321 GCGCGAAGGGTTCGCTAGA 1340

RESULT 2

AF191501

LOCUS

DEFINITION

AP191501

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

CDS

1..1586

/organism="Mus musculus"

/db_xref="taxon:10090"

/tissue_type="skeletal muscle"

1..1586

/gene="Hrnbpl1"

166..1419

/gene="Hrnbpl1"

/codon_start=1

/product="hexaribonucleotide binding protein 1"

/protein_id="AA183425.1"

/db_xref="GI:19032414"

/translation="MLASQVLLHSYGVPMIVPAAPVFPGLMQGNOEAAAAADPTMAQ

VASQFAPQNGIPAEYTAHPHPAPETGTTGTTVPDHTLNLYPTQTHSEASDTSQ

TVSGTATDDAETDQGFQPSSENTESQPKRLHVSNIIPFRDLDLRFQFG

KILDVEIIFRSGKGFQTFENSADAREKLHGTVEGRKIENVNATFVMTNK

KVNPYTNKLNPNVCAVYSPDFYAGTVLLCOANQSGSSMYSGPSLVYTSAMPFP

YPAASAAAAYRGALHRLGRGTGYNTLRAAAPPPPIPAYGVVYQEPVYGNKLOGVA

AVRYAQPPTATAAAYSDSYGRVYAADPVHHTLAPATYGVGMNAFAPLTDAKTRSHA

DVGLVLSLQASIVRGYNRPAY"

BASE COUNT 407 a 458 c 374 g 347 t

AP191501 1586 bp mRNA linear ROD 01-MAR-2002
Mus musculus hexaribonucleotide binding protein 1 (Hrnbpl1) mRNA,
complete cds.
AP191501
AP191501.1 GI:19032413
Mus musculus.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Chen, W. and Winkelman, J. C.
Direct Submission
Submitted (01-OCT-1999) Internal Medicine/Hematology-Oncology,
University of Cincinnati College of Medicine, 231 Bethesda Ave.,
Cincinnati, OH 45267-0508, USA

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2003, 10:46:57 ; Search time 3583.42 Seconds
(without alignments)
10882.841 Million cell updates/sec

Title: US-09-809-545A-1
Perfect score: 1340
Sequence: 1 gggcgccctgacacaat.....ggcgagggttcgctaga 1340

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1340	100.0	1340	6	AX268800	AX268800 Sequence
2	1145.8	85.5	1586	10	AF191501	AF191501 Mus muscu
3	1110	82.8	2002	10	AB041596	AB041596 Mus muscu
4	1084.4	80.9	1363	10	AF107204	AF107204 Mus muscu
5	989.4	72.3	2000	9	AF229057	AF229057 Homo sapi
6	906.4	67.6	3348	9	AF109106	AF109106 Homo sapi
7	905.2	67.6	1513	9	AK001027	AK001027 Homo sapi
8	904.2	67.5	2279	9	AF107203	AF107203 Homo sapi
9	861.8	64.3	1475	9	AB060859	AB060859 Macaca fa
10	857.2	64.0	2372	6	AF134676	AF134676 Sequence
11	796.4	59.6	1547	9	AF094849	AF094849 Homo sapi
12	541.2	40.4	1623	9	HSM803066	AF387322 Mus muscu
13	295.2	22.0	1134	10	AF387322	AF387322 Mus muscu
14	295.2	22.0	1695	10	AF229055	AF229055 Mus muscu
15	285.4	21.3	115860	2	AC120661	AC120661 Rattus no
16	281.2	21.0	1721	10	BC027263	BC027263 Mus muscu
17	280	20.9	1876	9	HSP38A20	AL009266 H. sapien
18	277.6	20.7	1925	9	AF229058	AF229058 Homo sapi
19	276	20.6	1555	9	BC025281	BC025281 Homo sapi
20	275.6	20.6	1393	9	AY072786	AY072786 Homo sapi
21	274	20.4	1558	9	AK055213	AK055213 Homo sapi
22	195.8	14.6	959	10	AF229056	AF229056 Mus muscu
23	186	13.9	1538	9	BC013115	BC013115 Homo sapi
24	172	12.8	2971	9	HSFOX14	AF109120 Homo sapi
25	172	12.8	164538	9	AC005774	AC005774 Homo sapi
26	145.8	10.9	2457	3	AY118678	AY118678 Drosophil
27	133.8	10.0	169192	2	AC114672	AC114672 Mus muscu
28	133.8	10.0	235183	2	AC079427	AC079427 Mus muscu
29	123.8	9.2	629	9	HSFOX03	AF109109 Homo sapi
30	123.8	9.2	112976	9	AC027683	AC027683 Homo sapi
31	123.8	9.2	169765	9	AC006075	AC006075 Homo sapi
32	120.6	9.0	105907	2	AC026429	AC026429 Homo sapi
33	111	8.3	251	9	HSFOX09	AF109115 Homo sapi
34	109.6	8.2	1426	3	CEU14946	U14946 Caenorhabdi
35	94.2	7.0	219	9	HSFOX05	AF109111 Homo sapi
36	90.4	6.7	470	9	HSFOX02	AF109108 Homo sapi
37	89.2	6.7	203649	10	AL603843	AL603843 Mouse DNA
38	88.6	6.6	108130	2	AC098922	AC098922 Rattus no
39	88.6	6.6	156820	2	AC123391	AC123391 Rattus no
40	87.4	6.5	596	9	HS323688	AJ323688 Homo sapi
41	87.4	6.5	617	9	HS335811	AJ335811 Homo sapi
42	87.4	6.5	653	9	HS335616	AJ335616 Homo sapi
43	87.4	6.5	675	9	HS334013	AJ334013 Homo sapi
44	87	6.5	108315	9	HS41P2	AL049748 Human DNA
45	81.8	6.1	374	9	HSFOX08	AF109114 Homo sapi

ALIGNMENTS

RESULT 1	AX268800	AX268800	1340 bp	DNA	linear	PAT 29-OCT-2001
LOCUS	AX268800	Sequence 1 from Patent WO0174901.				
DEFINITION	AX268800	Sequence 1 from Patent WO0174901.				
ACCESSION	AX268800					
VERSION	AX268800.1	GI:16541860				
KEYWORDS						
SOURCE		Norway rat.				
ORGANISM		Rattus norvegicus				
		Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;				
		Rattus.				
REFERENCE	1	Stanton,L.W. and White,R.T.				
AUTHORS		Secreted factors				
TITLE						